

Thomas Stäcker

Das VD17 at your fingertips: Der Masterplan. Nachgedanken zu einem paradigmatischen Digitalisierungsprogramm

Das Verzeichnis der im deutschen Sprachraum erschienenen Drucke des 17. Jahrhunderts¹ kann als eine Erfolgsgeschichte deutscher Forschungsförderung verbucht werden.² Mit der nahezu vollständigen Verzeichnung aller Drucke dieses für die Geschichte und Entwicklung Deutschlands so wichtigen Jahrhunderts ist eine Forschungsgrundlage geschaffen worden, die im europäischen und internationalen Zusammenhang ihresgleichen sucht. Es ist nicht nur die hohe Zahl an erschlossenen Altbeständen, die überzeugt, es ist auch das außerordentlich hohe Niveau der Erschließung und Konsistenz der Erfassung, die Zeichen gesetzt hat. Dem Projekt kam zugute, dass seine Konzeption anders als für den Vorgänger, das VD16, in eine Zeit fiel, in der sich die Datenbank bereits als festes Instrument der bibliothekarischen Erschließung etabliert hatte und die gedruckte Bibliografie im Niedergang begriffen war. Zugute kam dem Projekt weiterhin, dass es erste Schritte der Digitalisierung integrieren konnte, so dass heute den Drucken alle Schlüsselseiten in digitaler Form beigelegt sind; viele von ihnen bildeten im Falle kleinerer Werke bereits damals das gesamte Werk in digitaler Form ab (wenn auch noch in Schwarz-Weiß). Schon während der über zehnjährigen Laufzeit des Projekts wurden Stimmen laut, der Katalogisierung eine Digitalisierung folgen zu lassen.

Die Bemühungen um die Digitalisierung des deutschen schriftlichen Kulturguts reichen tatsächlich weit zurück und es gab wiederholt Anläufe, es Gestalt werden zu lassen. Der erste konzertierte Versuch bestand nach tastenden Versu-

¹ Vgl. <http://www.vd17.de>. Alle aufgeführten Internetressourcen wurden zuletzt am 22. Dezember 2017 aufgerufen.

² Vgl. u. a. Feuerstein-Herz, Petra: Neue Wege im VD 17 – Stand und Perspektiven in der nationalbibliografischen Verzeichnung. In: *Bibliothek – Forschung und Praxis* 35 (2011) H. 3. S. 411–414; Fabian, Claudia, Thomas Stäcker u. Annette Wehmeyer: Fortführung des VD 17. Perspektiven und Angebote der Trägerbibliotheken. In: *Bibliothek und Wissenschaft* 43 (2010). S. 193–197; Bürger, Thomas u. Ines Wolf: Die Schmelze des barocken Eisberges. Das Verzeichnis der deutschen Drucke des 17. Jahrhunderts (VD17). In: *BIS – Das Magazin der Bibliotheken in Sachsen* 2 (2009) H. 3. S. 162–166; Stäcker, Thomas: VD 17 – Mehr als eine Zwischenbilanz. In: *Zeitschrift für Bibliothekswesen und Bibliographie* 51 (2004) H. 4. 213–221.

chen in den so genannten Massendigitalisierungsprojekten³ der Deutschen Forschungsgemeinschaft (DFG). Dem vorausgegangen war ein Antrag der Sammlung Deutscher Drucke (SDD) an die DFG zur Digitalisierung der im Rahmen der SDD erworbenen Drucke. Die Verortung bei der SDD machte seinerzeit durchaus Sinn, denn die SDD bot als verteilte digitale Nationalbibliothek einen attraktiven Rahmen für ein solches Projekt und besaß mit den dort vertretenen Bibliotheken ein leistungsstarkes Konsortium. Bedauerlicherweise konnte die DFG sich zu dieser Zeit nicht dazu durchringen, das für damalige Verhältnisse sehr ehrgeizige Vorhaben zu fördern. Mit dem Google-Vertrag der Bayerischen Staatsbibliothek München (BSB) 2007 änderten sich dann die äußeren Rahmenbedingungen und machten eine Weiterführung der Überlegungen, ein koordiniertes Projekt allein aus öffentlichen Mitteln zu organisieren, hinfällig. An dessen Stelle traten Einzelprojekte, die sich um das Google-Projekt der BSB gruppierten und Lücken, die Google voraussichtlich lassen würde, zu füllen suchten. Als Orientierungsrahmen dienten die VDs, in die die BSB auch die Google-Digitalisate eintrug, wodurch Zug um Zug auf der Grundlage der nationalbibliografischen Verzeichnung eine nationale digitale Bibliothek entstand. Dabei wählte man im VD16 einen akkumulierenden Weg, indem Digitalisate ergänzt wurden, die nicht durch die BSB oder den jeweils nächsten Einlieferer, insbesondere die ULB Halle, eingebracht worden waren, um Doppeldigitalisierungen zu vermeiden. Im VD17 ging man von Anfang an einen anderen Weg, weil sich schnell abzeichnete, dass die ca. dreifache Menge nicht durch eine Bibliothek allein oder durch serielle Abarbeitung in definierter Zeit hätte digitalisiert werden können. Dies war die Geburtsstunde des Masterplans zur Digitalisierung der im VD17 nachgewiesenen Drucke.⁴ Der Masterplan war in gewisser Weise ein Novum in der deutschen Bibliothekslandschaft, da hier, vergleichbar allenfalls mit dem langjährigen Handschriftenbeschreibungsprogramm der DFG, versucht wurde, die Digitalisierung nicht mehr nur in mehr oder weniger zufällig initiierten Einzelprojekten, typischerweise von drei bis fünf Jahren Laufzeit, anzugehen, sondern programmatisch in einem festgelegten mehrjährigen Ablauf. Die Federführung des Vorhabens übernahm die Herzog August Bibliothek in Wolfenbüttel und sorgte weitgehend mit eigenen Ressourcen für die Koordinierung.

³ Vgl. Opitz, Andrea u. Thomas Stäcker: Workshop der Massendigitalisierungsprojekte der Deutschen Forschungsgemeinschaft an der Herzog August Bibliothek. In: *Zeitschrift für Bibliotheks- und Bibliographie* 56 (2009) H. 6. S. 363–373.

⁴ Vgl. Masterplan zur Digitalisierung des VD 17. Wolfenbüttel 2010. <http://www.vd17.de/files/vd17-masterplan-1.pdf>.

Der Masterplan musste verschiedene Probleme lösen. Die wichtigste Frage betraf die Aufteilung der prospektiven Gesamtmenge von nahezu 300 000 Titeln des VD17 unter den Einrichtungen, die sich zur Teilnahme am Masterplan bereit erklärt hatten, und natürlich die Vermeidung von Dubletten. Dabei konnte man sich nicht nur auf die Datenbank als Referenzinstrument stützen. Wollte der Masterplan zu einem erfolgreichen Konzept werden, das möglichst viele Einrichtungen nutzten, durfte er nicht auf die Institutionen beschränkt bleiben, die Teilnehmerbibliotheken des VD17 gewesen sind. Auch die scheinbar plausible Idee, sich auf Unika zu konzentrieren, um Dubletten zu vermeiden, hatte den entscheidenden Nachteil, dass dadurch Titel aus dem Raster fielen, die besonders verbreitet und an mehreren Bibliotheken vorhanden waren, was der Intention des Programms, von Anfang an ein möglichst breites Bild der Publikationen des 17. Jahrhunderts anzubieten, zuwidergelaufen wäre. Man verständigte sich schließlich auf vier Grundtypen, die alle Aspekte berücksichtigten: „Typ 1 beinhaltet Projekte mit nach Maßgabe des VD 17 unikalem Bestand, Typ 2 mit regional definiertem Bestand, Typ 3 Projekte, die zum Ziel haben, die ‚Mainstream-Literatur‘ des VD 17 zu digitalisieren, Typ 4 Projekte, die auf der Grundlage einer Liste von gezielt selektierten VD 17-Titeln vorgehen.“⁵

Wichtige Voraussetzung für die Durchführung des VD17-Masterplans waren Pilotprojekte wie die oben genannten Massendigitalisierungsprojekte der DFG, deren Zahlen heute eher bescheiden anmuten, die aber großen Einfluss auf die Gestaltung der späteren Workflows hatten. Besonders die Frage der Scanroboter wurde immer wieder intensiv diskutiert. Es zeigte sich im Verlauf der Vorhaben jedoch, dass die Roboter ihre Vorteile angesichts des materiell sehr heterogenen Materials nicht ausspielen konnten. Unterschiedliche Papierdicken und enge Heftungen, Verstaubungen oder auch Beschädigungen ließen kaum eine serielle Arbeit zu, sodass durch den Einsatz von Robotern anders als für spätere Jahrhunderte im 17. Jahrhundert kaum Geschwindigkeitsgewinne zu erzielen waren.

Nicht minder wichtig als Voraussetzung waren die Praxisregeln der DFG zur Digitalisierung.⁶ Deren Bedeutung kann kaum überschätzt werden. Sie sorgten dafür, dass in Projekten, die von der DFG gefördert wurden, die Scans einheitlich und von hoher Qualität waren. Zugleich wurden infrastrukturelle Rahmenbedingungen zur Erschließung formuliert, wie die auf METS/MODS basierende OAI-Schnittstelle, die Grundlage für die Nutzung des so genannten DFG-Viewers wurde, dessen Nutzung für das VD17 verpflichtend vorgeschrieben

⁵ Masterplan (wie Anm. 4), S. 1.

⁶ Vgl. Deutsche Forschungsgemeinschaft: Praxisregeln „Digitalisierung“. Bonn 2016. http://www.dfg.de/formulare/12_151/12_151_de.pdf.

wurde. Es ist heute gerade diese Schnittstelle, die großes Potenzial bietet und auf der auch weitere Services aufgebaut werden können. So wurde z. B. im Rahmen des OCR-D-Projekts diskutiert, ob man die Schnittstelle nutzen könne, um Bilder für OCR-Prozesse zu harvesten. In späteren Phasen des Projekts traten dann noch Forderungen nach Open Access und freien Lizenzen hinzu, die eine wichtige Voraussetzung für Nachnutzungen im Bereich der Digital Humanities bilden.

Ausgestattet mit dieser Erfahrung und gemeinsamen technischen Rahmenbedingungen konnte der Masterplan an den Start gehen. Die teilnehmenden oder sich für den Masterplan interessierenden Bibliotheken konnten auf dieser Grundlage Einzelanträge stellen, indem sie sich zur Einhaltung der dort formulierten Regeln verpflichteten. Das Besondere daran war, dass einerseits ein einheitlicher Rahmen für alle Projekte im Rahmen des Masterplans gegeben war, zugleich aber nach wie vor die Antragsmechanismen der DFG wirksam blieben, die auf Einzelprojektförderung zielten. So konnte man sich zwar hinsichtlich der Erläuterung der Antragsziele kurzfassen und auf den Masterplan verweisen, zugleich musste aber jeder einzelne Antrag das eigene Vorhaben nach Maßgabe der Bestände und jeweiligen Infrastruktur eigens begründen und die gutachterlichen Hürden überwinden. Hier setzte wiederholt Kritik, gerade auch der Förderer, an, die angesichts der hier auftretenden Kostenschwankungen der Projekte zur besseren Kalkulierbarkeit der Vorhaben einheitlichere Preise für die Digitalisierung forderten, ein Anliegen, das später im VD18 sehr zum Schaden des Vorhabens umgesetzt wurde. Problematisch an den dann im VD18 durchgesetzten Einheitspreisen war dabei weniger die Tatsache, dass überhaupt Richtwerte eingeführt, sondern wie sie festgesetzt wurden. Denn die Zahlen, die hier kalkuliert wurden, basierten auf problematischen Grundannahmen, die wichtige Faktoren wie Titelmenge und infrastrukturelle Rahmenbedingungen außer Acht ließen, sodass heute VD18-Projekte, die anders als die Pilotprojekte, die sich zur raschen Erzielung von Mengenvorgaben auf umfangreichere Titel konzentriert hatten, bei gleicher Seitenmenge mit höheren Titelzahlen kämpfen müssen und mit dem knappen Seitenpreis nicht auskömmlich digitalisieren können. Leider wurde im Rahmen des VD17 versäumt, sich dieses Problems anzunehmen und belastbare Formeln zu entwickeln, die die echten Kosten hätten adäquat abbilden können. Grundlage einer solchen Formel sind einerseits gut messbare Größen wie die Anzahl der Seiten und der Titel, andererseits Aspekte des Bestands, der gegebenenfalls schwierigen Aushebung, der konservatorischen Rahmenbedingungen oder auch des vorhandenen Equipments. Letztere sind für außenstehende Gutachter oft schwer zu durchschauen. Trotzdem ist es den Antragstellern des VD17-Masterplans meist gut gelungen, sie plausibel zu machen und von der Norm abweichende höhere Kosten nachvollziehbar zu be-

gründen. Für die Finanzierung dieses Segments hätte man einen Standardwert ansetzen können, der, wenn er nicht auskömmlich ist, eigens gerechtfertigt werden muss, allerdings nur bis zu einer bestimmten Obergrenze. So hätte sich eine tragfähige Formel des Zeitaufwands ergeben: Rüstzeit pro Titel + Zeit zur Digitalisierung der Seite, wobei die Rüstzeit bei einem normativen Wert für T_{\min} keiner eigenen Erklärung bedürfte, bei einer Abweichung bis T_{\max} aber jeweils eine den Geschäftsgang und Besonderheit des Bestands ausführlich beschreibende Begründung gegeben werden müsste. Die Zeit der Digitalisierung der Seite hätte man normativ vorschreiben können, da hier relevante Equipmentfragen Sache der Optimierung des Produktionsprozesses sind. Da solche Überlegungen im Rahmen des VD17 nicht weiterverfolgt wurden, blieben die Kostenstrukturen nahezu gänzlich unbestimmt und führten zugleich zu einer schweren Hypothek für das VD18.

Weitgehend erfolgreich verlief die Organisation der Projekte und die Koordination unter den zahlreichen Partnerbibliotheken. Beteiligt waren bisher die Bibliotheken UB Augsburg, SB Berlin, SLUB Dresden, UB Erlangen-Nürnberg, UB Frankfurt am Main, SUB Göttingen, FB Gotha, ULB Halle, SUB Hamburg, ThULB Jena, LBZ Koblenz, StUB Köln, UB Leipzig, StB Nürnberg, UB Rostock, HAAB Weimar und die HAB Wolfenbüttel. Die Beiträge schwankten nach Leistungsfähigkeit und Bestandsdichte zwischen 1500 und 15 000 Titeln pro Projekt. Um den operativen Aufwand zu begrenzen und den Charakter des Masterplans als Massendigitalisierungsprojekt nicht preiszugeben, war eine Mindestdigitalisierungsmenge von 1500 Titeln anzustreben. Kalkulatorische Schwierigkeiten gab es dank der guten bibliografischen Vorarbeiten selten, allein die Projekte des Regionaltyps, die sich nicht auf vorhandene VD17-Katalogisate stützen konnten, hatten hin und wieder mit zu optimistisch geschätzten Titelmengen zu kämpfen, ohne dass das Anliegen grundsätzlich gefährdet worden wäre. Einige wichen pragmatisch auf die Digitalisierung nach Nummernliste, sozusagen nach dem pick-and-choose-Modell aus. Die Partner legten in Abstimmung mit der Koordinierungsstelle in Wolfenbüttel das jeweilige Projekt und die zu digitalisierenden Volumina fest, ohne dass eine Verpflichtung dazu bestanden hätte. Im Rückblick erwies sich allerdings der Umstand, dass Wolfenbüttel die Koordinierungsarbeiten zwar mit Unterstützung der Gruppe der Partnerbibliotheken, aber doch fast vollständig aus eigenen Ressourcen übernommen hatte, als Achillesferse, denn immer wieder belastete die sonstige Tagesarbeit die Projektkoordination. Man hätte weit effizienter wirken können, wenn die jeweiligen Projekte hätten enger begleitet und bei der Vorbereitung intensiver unterstützt werden können. Zudem lagen auch Schulungen und Einführungen an, die Wolfenbüttel weitgehend als Eigenleistung erbrachte. So blieb das Vorhaben gewissermaßen unter seinen Möglichkeiten, die nur mit

einer zusätzlichen Kraft zur zentralen Koordinierung und Steuerung hätten voll ausgeschöpft werden können. Neue programmatische Digitalisierungsvorhaben sollten daher grundsätzlich mindestens eine halbe, besser und nach Größenordnung des Projekts und nach Anzahl teilnehmender Partner eine ganze wissenschaftliche Koordinierungsstelle vorsehen, die konzeptionell und organisatorisch die Maßnahmen begleitet, Zeitpläne abstimmt, Lücken identifiziert, neue Vorhaben initiiert, Regel- und Schnittstellenkonformität sicherstellt (z. B. DFG-Viewerkonformität) und die Rahmenbedingungen prüft. Dennoch hat das VD17-Projekt in relativ kurzer Zeit mit einem vergleichsweise beschränkten Budget bemerkenswerte Ergebnisse erzielt. Mit Stand von Ende Dezember 2017 sind knapp 155 000 Digitalisate online zugänglich. Freilich sind 65 000 davon der Google-Initiative zuzurechnen, ohne die der Prozess weit länger gedauert hätte. Umgekehrt sind die Digitalisate, die mit öffentlichen Mitteln entstanden sind, theoretisch ohne Einschränkung frei und können von den besitzenden Einrichtungen z. B. unter offenen CC-Lizenzen zur Verfügung gestellt werden, sodass eine Nachnutzung zum Zwecke der Publikation oder Aggregation der Digitalisate an anderen Orten ungehindert möglich ist. In diesem Sinne hat die Google-Initiative, deren ersten Phase 2016 zu Ende ging,⁷ viel Gutes bewirkt, weil es der öffentlichen Hand kaum gelungen wäre, in so kurzer Zeit diese Mengen zu digitalisieren. Zugleich ist aber sichtbar, dass Fragen bleiben, etwa bei der Qualität der Scans (z. B. Problem der nicht mitgescannten Faltafeln) oder den Nachnutzungsrechten. Die nähere Zukunft wird zeigen, ob und in welchem Umfang die weitere Verbesserung des Korpus und die Nutzung etwa im Kontext von OCR-D⁸ ermöglicht wird. Dass die öffentliche Hand bzw. die Gesellschaft in uneingeschränktem Besitz ihres Kulturguts bleiben muss, gleichgültig, ob es in analoger oder digitaler Form vorliegt, ist eine unverzichtbare Forderung an alle Digitalisierungsprojekte, die in einer Public-private-Partnership entstehen, und es ist gut zu wissen, dass wohl weitgehend alle Einrichtungen, die mit Google ko-

7 Eine offizielle Nachricht dazu findet sich leider nicht (vgl. <https://www.bsb-muenchen.de/ueber-uns/kooperationen/google/>), nur die Aussage des Direktors (2017), dass die Kampagne „sang- und klanglos im letzten Jahr“ zu Ende ging: <https://www.golem.de/news/google-books-die-bibliotheken-leben-noch-1705-127647.html>. Auf Nachfrage erhielt ich von der BSB München die Erläuterung, dass das Hauptziel, also das Digitalisieren von mehr als einer Million Bände aus dem Bestand der Bayerischen Staatsbibliothek, tatsächlich bereits seit über einem Jahr erreicht ist. Da aber das Projekt Google Books weiterläuft und die BSB Partner in diesem Projekt ist und bleibt, können jährlich ca. 15 000 bis 20 000 weitere Bände digitalisiert werden, sodass das vereinbarte Projekt abgeschlossen ist, zugleich aber das Digitalisieren weitergeht.

8 <http://www.ocr-d.de>.

operieren, und so auch die BSB, alle Scans wirklich als File besitzen.⁹ Von der Koordinierungsseite her war das Google-Projekt der BSB insofern schwierig, als man sich vorab weder auf gemeinsame Qualitätsstandards noch auf ein koordiniertes Vorgehen verständigen konnte. Das Konzept der Bestandsdigitalisierung einer Sammlung ließ z. B. Fragen des besten Exemplars nicht zu und eine ausgewogene Balancierung der Workloads für eine verteilte Kampagne, verbunden auch mit einer Berücksichtigung der Interessenlagen einzelner Bibliotheken etwa an der Digitalisierung von Druckorten, konnte so nicht oder nur mit der Umständlichkeit einer Bestandsprüfung gegen die BSB unternommen werden. So gesehen bestand der Masterplan nicht nur aus einem Projekt, sondern aus zweien, die unterschiedliche Wege einschlugen und Methoden nutzten. Trotz dieser Schwierigkeiten konnten aber die meisten dieser Fragen, nicht zuletzt durch die umfassende Zu- und Mitarbeit der bayerischen Kollegen, geklärt werden und über das VD17 entstand ein Nachweisinstrument zu einem Korpus der im deutschsprachigen Raum des 17. Jahrhunderts erschienenen Literatur, das beide Projektklinien gut integriert.

Mit Stand vom 17. Dezember 2017 sind im Projekt 152 853 Digitalisate nachgewiesen.¹⁰ Nimmt man die derzeit bewilligten Projekte dazu, werden aus der jetzigen Kampagne 220 000 Drucke online zur Verfügung stehen.¹¹ 65 000 Drucke stammen aus der Google-Digitalisierung, also momentan 30 %. Vorgesehen ist nach Masterplan eigentlich, dass 95 % des VD17 in zwei Phasen digitalisiert werden,¹² der Rest nach Bedarf in kleineren Digitalisierungsprojekten. Seinerzeit (2010) war man allerdings noch von einer Titelmenge von ca. 285 000 ausgegangen, wenngleich mit der Erwartung, dass die Titelmenge ansteigt. Diese Erwartung hat sich erfreulicherweise erfüllt. Derzeit sind es nahezu 300 000 Ausgaben (299 733; Stand vom 17. Dezember 2017). Der genaue Zuwachs bzw. das Gesamtvolumen aller Titel ist schwer zu kalkulieren, weil immer noch große Sammlungen, nicht nur, aber vor allem in Osteuropa fehlen, sodass man wahrscheinlich auf absehbare Zeit keine belastbare Größenordnung der Gesamttitelproduktion im 17. Jahrhundert bekommen wird.¹³ Die Müller-Stu-

9 Ceynowa auf golem.de: „Den Datenbestand hostet die BSB selbst. ‚Es war uns wichtig, dass wir die komplette Verfügungsgewalt über die Digitalisate haben.‘“ <https://www.golem.de/news/google-books-die-bibliotheken-leben-noch-1705-127647.html>.

10 Vgl. <http://gso.gbv.de/DB=1.28/CMD?ACT=SRCHA&TRM=onx+34+erj+1601-1700>.

11 Nach Angaben der HAB Wolfenbüttel.

12 Masterplan (wie Anm. 4), S. 6.

13 Z. B. sind Vergleiche mit Daten in den von Garber herausgegebenen Personalschriften sehr ernüchternd. Manche Orte haben Deckungsraten von weniger als 20 %. Vgl. Klaus Garber (Hrsg.): Handbuch des personalen Gelegenheitsschrifttums in europäischen Bibliotheken und Archiven. Hildesheim u. a.: Olms-Weidmann 2001–2013.

die, die seinerzeit zur Berechnung des Umfangs angefertigt worden war, hat die Umfänge mit 265 000 Titeln drastisch unterschätzt.¹⁴ Geht man aber von einer pragmatischen Größenordnung von 325 000 Drucken aus und reduziert den Digitalisierungsgrad auf zuverlässig in deutschen Bibliotheken erreichbare 90 %, ¹⁵ wäre eine Zielgröße für den Masterplan 292 500 Drucke. Es fehlen also derzeit noch 72 500 Drucke, was in etwa dem projektierten Volumen der zweiten Phase entspricht. Danach werden, sofern die DFG das Digitalisierungsprogramm konsequent fortsetzt, die meisten der im VD17 erschienenen Drucke digitalisiert sein. Es stünde dann ein im europäischen Kontext einzigartiges und unschätzbar wertvolles Reservoir für die Erforschung des 17. Jahrhunderts zur Verfügung, das es nicht nur erlaubt, bequem auf die sonst oft schwer zugänglichen Titel zuzugreifen, sondern das auch die Grundlage für weitere Digitalisierungskampagnen zur Herstellung von maschinenlesbaren Fassungen per OCR (vgl. oben OCR-D), z. B. für DH-affine Forschungsfragen, bildet.

Der Masterplan zum VD17 gehört neben dem späteren Masterplan zur Zeitungsdigitalisierung,¹⁶ der noch nicht in der Breite angelaufen ist, zu den erfolgreichsten abgestimmten Digitalisierungskampagnen und hat in seiner Bedeutung und Tragweite als produktives Digitalisierungskonsortium andere Initiativen deutlich hinter sich gelassen. Der Masterplan ist in seiner Genese eine merkwürdige, aber eben auch typisch deutsche Initiative. Seine heute mehr und mehr sichtbare Effizienz spiegelt sich nicht in wissenschaftspolitischen PR-Strategien oder spektakulären politischen Auftritten, ja der Prozess ging weit hin ohne Aufheben voran. Wie ein zuverlässiges Uhrwerk arbeiteten die beteiligten Bibliotheken ihr Programm ab, mit bibliothekarischer Effizienz und mit dem ihnen so eigenen Sinn für Pragmatik, ohne dass es komplexer Begleitstudien oder Beraterstäbe bedurft hätte. Es lief und läuft so geräuschlos ab, dass die Bundesregierung in ihrem Legislaturbericht zur Digitalen Agenda 2014–2017¹⁷ noch nicht einmal Notiz von den Aktivitäten nahm, die sich doch immerhin darauf richteten, ein ganzes Jahrhundert deutscher Literatur zu digitalisie-

14 Müller, Wolfgang: Die Drucke des 17. Jahrhunderts im deutschen Sprachraum. Untersuchungen zu ihrer Verzeichnung in einem VD17. Wiesbaden: Harrassowitz 1990.

15 Ein großes Problem ist, dass die DFG mit Ausnahme Israels ausländische Projekte nicht fördert. Versuche des Verfassers, kooperative Projekte mit anderen Förderern, insbesondere mit Polen, zu initiieren, scheiterten bedauerlicherweise an den finanziellen Möglichkeiten, aber auch schwierigen Förderstrukturen der polnischen Seite. Europäische Programme zur Digitalisierung, die hier hätten greifen können, fehlen leider und sind nach wie vor ein dringendes Desiderat.

16 Vgl. http://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/erschliessung_digitalisierung/index.html.

17 Vgl. Bundesministerium des Innern, Bundesministerium für Wirtschaft und Energie u. Bundesministerium für Verkehr und digitale Infrastruktur (Hrsg.): Legislaturbericht Digitale Agen-

ren. Und das, obwohl die Agenda behauptet hatte: „Ferner treiben wir die Digitalisierung von Kulturgut weiter voran und verbessern die Zugänglichkeit zum kulturellen und wissenschaftlichen Erbe in Archiven, Bibliotheken und Museen“.¹⁸ Tatsächlich hat man trotz der national strategischen Ausrichtung der Agenda nicht versucht, Konzepte zu einer Gesamtdigitalisierung des deutschen Kulturerbes zu entwickeln. Das mag zwar dem in diesem Bereich wirksamen Föderalismusproblem geschuldet sein, allerdings ist dessen Fehlen doch erstaunlich, weil parallel dazu, wenngleich holperig, die Etablierung der DDB gelang, die trotz der stupenden Menge an Nachweisen einerseits nur einen Bruchteil des deutschen Kulturguts nachweist, andererseits auch keine steuernden Funktionen in der Frage der Kulturgutdigitalisierung übernommen hat. Selbst auf der europäischen Ebene blieb und bleibt es still. Der Masterplan verstand sich mit der Fokussierung auf das Verzeichnis der im deutschen Sprachraum erschienenen Drucke immer auch als Teilprojekt eines gesamteuropäischen Unternehmens (vgl. z. B. „Gallica“ für die französische Literatur). Doch während sich die europäischen Projekte zur Datenanalyse und Forschungsmethodik die Klinke in die Hand gaben, fristete die Retrodigitalisierung und digitale Produktion als offenbar „unsexy“ und reine Serienarbeit ein Nischendasein in europäischen Förderprojekten, obwohl doch allein die systematische digitale Transformation des europäischen Kulturerbes die Grundlage geschaffen hätte, die innovativen Tools überhaupt sinnvoll bei größeren Datenmengen einzusetzen.¹⁹ Bis heute gibt es keinen abgestimmten Plan zur Digitalisierung des europäischen Kulturguts, ja nicht einmal signifikant zu nennende, von der EU geförderte Digitalisierungsprojekte.

Vor diesem Hintergrund der wissenschaftspolitischen Vernachlässigung der Förderung des digitalen Rohstoffs kann es der DFG (vor allem dem LIS-Bereich) nicht hoch genug angerechnet werden, dass sie das Potenzial der Retrodigitalisierung als kulturbewahrenden und forschungsinitiiierenden Prozess erkannt und über Jahre systematisch gefördert hat. Dabei war auch dieser Prozess nicht frei von intrinsischen Schwierigkeiten, die bis heute andauern und sich vor allem um den Begriff der Forschungsrelevanz ranken, denn in der Bereit-

da 2014–2017. Frankfurt am Main, Mai 2017. https://www.digitale-agenda.de/Content/DE/_Anlagen/2017/04/2017-04-26-digitale-agenda.pdf?__blob=publicationFile&v=3.

18 Bundesministerium für Wirtschaft und Energie, Bundesministerium des Innern u. Bundesministerium für Verkehr und digitale Infrastruktur (Hrsg.): Digitale Agenda 2014–2017. München, August 2014, S. 29. https://www.bundesregierung.de/Content/DE/_Anlagen/2014/08/2014-08-20-digitale-agenda.pdf?__blob=publicationFile&v=6.

19 Vgl. das zu diesem Problemkomplex entstehende Papier des DARIAH-Stakeholdergremiums Wissenschaftliche Sammlungen „Agenda zur Kulturgutdigitalisierung – Wege durch den digitalen Dschungel“.

stellung digitaler Infrastruktur in Gestalt von Retrodigitalisaten bzw. Forschungsdaten liegt ein Henne-Ei-Problem. Forscherinnen und Forscher haben vor allem ihre konkrete Forschung vor Augen, die bestimmte Daten- und Dokumentbedarfe generiert. Diese sind fach- und disziplinspezifisch. Historiker brauchen andere Texte als Philologen, Kunsthistoriker andere als Philosophen. Aus Sicht der bereitstellenden Infrastruktur können all diese Spezialinteressen in der digitalen Dokument- und Datenbereitstellung nur bedingt und mit Blick auf eine möglichst breite Nachnutzbarkeit generisch befriedigt werden. Zugleich ist aber bekannt, dass Forschungsinteressen ebenso durch die Bereitstellung neuen Materials entstehen und Forschung selbst oft assoziierend nach dem Prinzip der Serendipity voranschreitet. Insofern sind die Forderungen nach dem „Was“ der Digitalisierung immer auch zirkulär. Forschung fragt nach Daten und digitalen Dokumenten, digitale Dokumente und Daten generieren Forschung. Ein anderes Argument, das gegen eine zu selektive Digitalisierung spricht, sind die höheren Kosten. Selbst wenn weite Teile des in Digitalisierungskampagnen konvertierten schriftlichen Kulturguts derzeit nicht beforcht werden, ist es doch weit günstiger, alles zu digitalisieren, als in kostenträchtigen, weil wissenschaftliche Expertise erfordernden Auswahlprozessen Subkorpora zu selektieren. Bei der Kulturgutdigitalisierung sollte die Kultur des pragmatischen „Alles“ herrschen und die Forschungsrelevanz dahingehend bestimmt werden, dass der Gesamtkörper der schriftlichen kulturellen Überlieferung, wenn er digital zur Verfügung stünde, auch alle potenziellen Forschungsfragen zu beantworten in der Lage ist. Dass dabei nicht Unerfüllbares verlangt wird, verdeutlichen der VD17-Masterplan und die begleitenden Vorhaben zum VD16 und VD18. Selbst wenn noch einige Jahre vergehen werden, ist die Komplettdigitalisierung des deutschen schriftlichen Kulturerbes keine Utopie, sondern ein konkretes operationalisierbares Ziel. Die Zahlen sind auf der Grundlage der ausgezeichneten bibliografischen Arbeiten früherer Generationen und plausiblen Schätzungen bzw. Studien hinreichend bekannt bzw. präzisierbar – in Deutschland aufbewahrte mittelalterliche Handschriften: 60 000, Gesamtzahl aller Inkunabelausgaben: 27 000,²⁰ im deutschsprachigen Raum erschienene Ausgaben des 16. Jahrhunderts: 120 000, Ausgaben des 17. Jahrhunderts: 325 000, Ausgaben des 18. Jahrhunderts: 600 000. Noch nicht ermittelt ist die Anzahl der Titel im 19. Jahrhundert, die auf mindestens drei Millionen geschätzt wird. Insofern ist es um die Ermittlung von Mengengerüsten der publizierten schriftlichen Kulturgüter bzw. der mittelalterlichen Handschriften

20 Hier ist eine Differenzierung nach „im deutschsprachigen Raum erschienen“, wie im 16. bis 19. Jahrhundert, oder „in Deutschland befindlich“, wie bei den mittelalterlichen Handschriften, aus pragmatischen Gründen nicht sinnvoll.

deutlich besser bestellt als etwa um die in deutschen Archiven aufbewahrten Handschriften und schriftlichen Nachlassmaterialien oder die in Museen aufbewahrten Sammlungsgegenstände.²¹ Es zeigt sich hieran auch, dass ohne vorbereitende und Daten zusammenführende Erschließungsarbeiten umfangreiche Digitalisierungsprogramme nicht oder nur partikular und selektiv möglich sind. Die Konsekutivität dieses Prozesses zeigt sich auch im jetzt gestarteten DFG-Programm zur Digitalisierung von in Archiven aufbewahrten Materialien, das unter der Voraussetzung des langjährigen Programms zur Erschließung der Findbücher steht.²² Katalogisierung und Erschließung bieten die unverzichtbaren Anker für Digitalisate. Mit Blick auf weitere Entwicklungen wie das Semantic Web sind diese Vorarbeiten, die die eindeutige Referenzierbarkeit sicherstellen, ebenso von entscheidender Bedeutung und werden in ihrer Notwendigkeit, gerade was historische Werke anbelangt, oft unterschätzt. Eine wichtige Aufgabe einer bibliografischen Beschreibung ist nach wie vor Eindeutigkeit. Das impliziert zweierlei: zum einen, dass die Einheit („Granularität“) des Referenzierungszieles bekannt ist, zum anderen, dass Merkmale vorhanden sind, die eine Einheit von der anderen zu unterscheiden erlauben. Beides ist Gegenstand eines Modellierungs- und auch Abstimmungs- bzw. Standardisierungsprozesses. So hat mit den FRBR²³ und dem von ihnen abgeleiteten RDA²⁴ eine wichtige Differenzierung stattgefunden, die das jahrhundertalte bibliothekarische „Ausgabenparadigma“ mit der Einführung des Werkbegriffs als eines zentralen Konzepts der Katalogisierung relativiert, eine entscheidende Änderung, die, wenngleich es schon in der Vergangenheit ähnliche Konzepte wie den „Einheitssachtitel“ gegeben hat, die Referenzierungs- und damit Identifizierungsmöglichkeiten deutlich erweitert und gerade für neue Entwicklungen wie die Nutzung von Wissensgraphen relevant werden könnte.

Will man eine Bilanz aus dem VD17-Digitalisierungsprogramm ziehen, so vor allem die, dass es auf der Grundlage der in Deutschland vorhandenen Ressourcen grundsätzlich möglich ist, ein auch umfangreiches nationales Digitalisierungsprogramm aufzusetzen. Wie so oft stehen koordinierten Programmen aber die föderalen staatlichen Strukturen im Wege, nicht in dem Sinne, dass es nicht Kooperationsmöglichkeiten gäbe, sondern eher, dass es schwer bis aus-

21 Siehe aber die bemerkenswerte Arbeit der Koordinierungsstelle für wissenschaftliche Universitäts-sammlungen (<http://wissenschaftliche-sammlungen.de/de/>), die zu mehr Transparenz im Feld der Universitäts-sammlungen geführt hat. Zahlen zu den insgesamt in Museen aufbewahrten Objekten sind nicht zu bekommen, vgl. das Papier des DARIAH-Stakeholdergremiums (wie Anm. 19).

22 Vgl. http://www.dfg.de/foerderung/info_wissenschaft/2016/info_wissenschaft_16_78/.

23 Vgl. <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>.

24 Vgl. <http://www.dnb.de/DE/Standardisierung/International/rdaFaq.html>.

sichtlos ist, zwischen den Ländern und dem Bund einen Konsens über gemeinsame Förderstrategien in diesem Feld zu erzielen. Zwar fördern die meisten Länder in mehr oder weniger umfangreicher Form die Kulturgutdigitalisierung, es gibt aber auf dieser Ebene so gut wie keine Anstrengungen, diese national zu verzahnen. Es ist daher fast einzig den Förderaktivitäten der DFG und in Teilen²⁵ denen des BMBF zu verdanken, dass auch übergreifende Digitalisierungskonzepte zum Tragen kommen. Man muss aus der Erfahrung leider hinzufügen, dass, will man zügig zum Ziel kommen, nur diese Art der Förderung sinnvoll ist. Maßnahmen der Länder haben bis heute diese Aktivitäten allenfalls flankiert; da aber nicht alle Länder gleichermaßen in der Lage und willens sind, Digitalisierungsprogramme zu fördern – meist in Verkennung des großen Potenzials, das solche Programme nicht nur für die Wissenschaft, sondern auch für die kulturelle Teilhabe der jeweiligen Bevölkerung und für die Attraktivität von Standorten haben –, bleiben solche Bemühungen Stückwerk oder zufällig. Will man daher tatsächlich zu einer flächendeckenden Digitalisierung des deutschen schriftlichen Kulturguts kommen, kann die Empfehlung nur lauten, dass Förderprogramme über die DFG oder das BMBF bzw. am besten in Abstimmung zwischen beiden großen nationalen Förderern konzipiert und forciert werden, und auch, dass man dort nicht auf Programme der Länder vertraut oder erwartet, dass die Länder sich in nationale Programme einbringen.²⁶ Quotierungen aus diesen Gründen (z. B. bei 60 %) führen in aller Regel nur zu Abbruch oder Zerfaserung des jeweiligen Programms. Sinnvoll ist indes, dass man Grundgesamtmengen ermittelt, von denen nach dem Paretoprinzip mindestens 80 % digitalisiert werden. Komplizierte Fälle oder die jeweiligen Reste, sind dann tatsächlich Ländersache und können bedarfsabhängig nachdigitalisiert werden. Priorisiert oder mit zusätzlichen Mitteln belohnt werden könnten aber Projekte, in denen sich Länder finanziell an der Digitalisierung beteiligen. Ländersache sind auch und vor allem regionale Materialien, die nur begrenzt von überregionalem und vor allem von landesgeschichtlichem Interesse sind, etwa die aus Pflichtablieferungen aufgebauten Bestände der regional sammelnden Landesbibliotheken.

Das VD17-Digitalisierungsprogramm hat, mit vergleichsweise geringem Mittelaufwand, die Perspektive für Projekte dieser Art geöffnet und den Weg für das VD18 und hoffentlich auch VD19 geebnet. Es zeigt aber ebenso, wie gleichsam generationenübergreifend die Räder ineinandergreifen. Nach insgesamt ca.

²⁵ Das BMBF hat in seinen Förderprogrammen den Fokus eher auf nichttextliche Materialien gelegt.

²⁶ Allerdings gibt es in jüngerer Zeit hier etwas Bewegung, sodass sich die Lage möglicherweise ändert.

15-jähriger Katalogisierung folgte die Phase einer 10- bis 15-jährigen Imagedigitalisierung, auf die jetzt eine noch nicht eindeutig einschätzbare Phase der Umwandlung in maschinenlesbare Texte folgt, an die sich dann, mit technischen Mitteln, ein immer weiter in die Strukturen der Texte vordringendes Analysieren anschließt. Die Entwicklung mahnt zu Geduld und langem Atem, aber auch zu Mut zu nationalen Strategien – für die Forschungspolitik und Förderer gleichermaßen. Es wird sich auszahlen.