

Findbücher ins Netz! Lösungen für kleine und mittelgroße Archive

von Andreas Pilger und Peter Worm

Im folgenden Beitrag soll es zunächst darum gehen, welche Hürden vor einer Online-Stellung von Findbüchern genommen werden müssen und welche praktischen Wege es gibt, diese Hürden zu meistern. Danach wird aufgezeigt, welche Funktionalitäten das Internetportal „Archive in NRW“ (www.archive.nrw.de) nach dem Relaunch zu Beginn dieses Jahres bietet und wie es mit Findbuchdaten bestückt werden kann.

und b) – fällt Folgendes auf: Die ländlichen Regionen sind überproportional stark vertreten. Gerade den Kreis Unna, den Märkischen Kreis und den Kreis Soest kann man als besonders aktiv hervorheben; seit 2009 ist zudem in den Kreisen Borken, Steinfurt, Minden-Lübbecke, Höxter und Olpe viel passiert. Umgekehrt zeigen die Großstadtarchive kaum Engagement oder präsentieren sich mit ihren Findmitteln nur auf der stadteigenen Homepage.

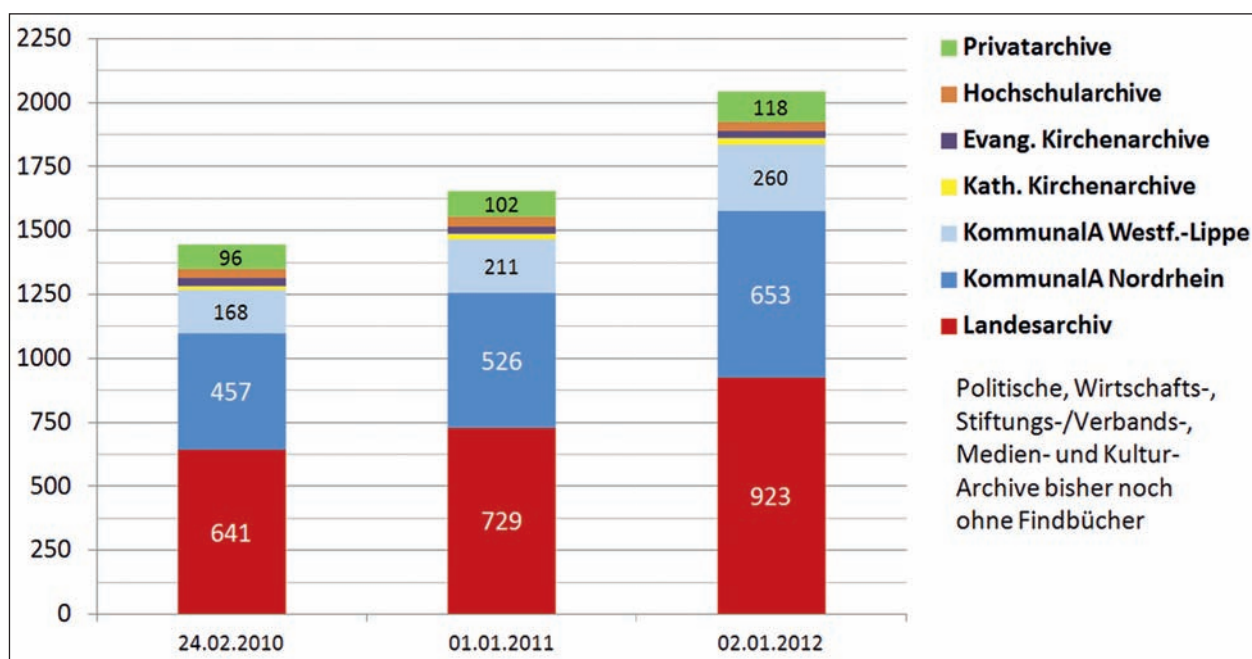


Abb. 1: Quantitative Entwicklung der Online-Findbücher in „Archive in NRW“

Statistische Betrachtung

Schaut man die Zahlen der in den letzten drei Jahren eingestellten Findbücher und Beständeübersichten im Archivportal „Archive in NRW“ an, so fällt auf, dass die Hälfte der Online Findbücher durch das Landesarchiv NRW eingestellt worden sind; ein Drittel der Online Findbücher stammt aus rheinischen Kommunen und nur ca. 12 % aus westfälischen Kommunalarchiven (vgl. Abb. 1).¹ Der Fairness halber muss man jedoch hinzufügen, dass rund zwei Drittel der rheinischen Findbücher dem Stadtarchiv Köln entstammen. Trotzdem haben wir erst in diesem Jahr die Marke überschritten, dass statistisch gesehen zumindest ein Findbuch eines jeden westfälischen Kommunalarchivs im Archivportal NRW online zur Verfügung steht. Das ist zu wenig, gerade im Hinblick auf die Vorüberlegungen, auch das Archivgut selbst über eine Förderlinie der Deutschen Forschungsgemeinschaft (DFG) online zu stellen.²

Vergleicht man die Verteilung der Aktivitäten im Land – hier in 2 Sichten von Ende 2009 und Anfang 2012 (Abb. 2a

Hinderungsgründe für eine größere Online-Präsenz

Die Gründe der Archive, warum der Schritt ins Web 1.0 nicht gewagt wird, können vielfältig sein. Im Folgenden soll auf drei Probleme eingegangen werden, die sich relativ leicht bewältigen lassen:

¹ Einen guten Überblick über die Entwicklung des Portals bieten: Mechthild Black-Veldtrup/Andreas Pilger/Martina Wiech, 11 Jahre archive.nrw.de – Bilanz und Perspektiven, in: Archivpflege in Westfalen-Lippe 71 (2009), S. 5–11; Peter Worm, Datenaustausch in der Praxis: Erfahrungen aus elf Jahren landesweitem, archivpartenübergreifendem Portal Archive in NRW, in: Katrin Wenzel/Jan Jäckel (Hrsg.), Retrokonversion, Austauschformate und Archivgutdigitalisierung. Beiträge zum 14. Archivwissenschaftlichen Kolloquium der Archivschule Marburg (Veröffentlichungen der Archivschule Marburg 51), Marburg 2010, S. 237–250; Andreas Pilger, Stand und Perspektiven des Portals „Archive in NRW“ nach dem Relaunch, in: Archivar 65 (2012), S. 208–209.

² Frank M. Bischoff und Marcus Stumpf, Digitalisierung von archivalischen Quellen – DFG-Rundgespräch diskutiert fachliche Eckpunkte und Ziele einer bundesweiten Digitalisierungskampagne, in: Archivar 64 (2011), S. 343–346.

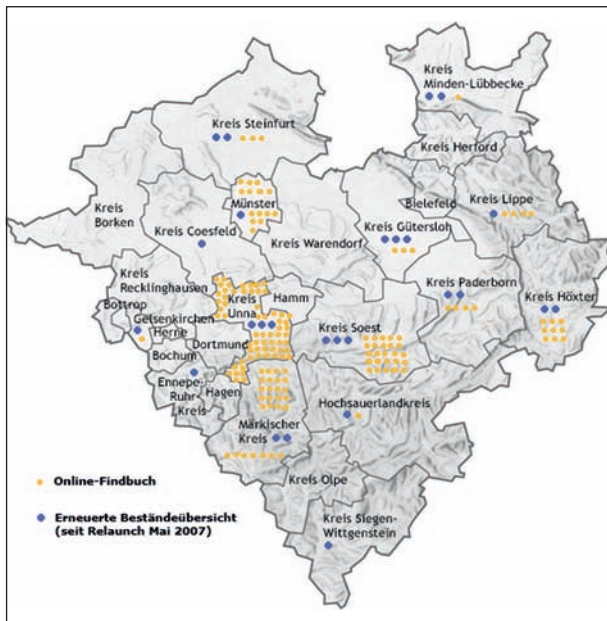


Abb. 2a: Westfälisch-Lippisches Engagement der Kommunalarchive bei „Archive in NRW“ (Stand November 2009)

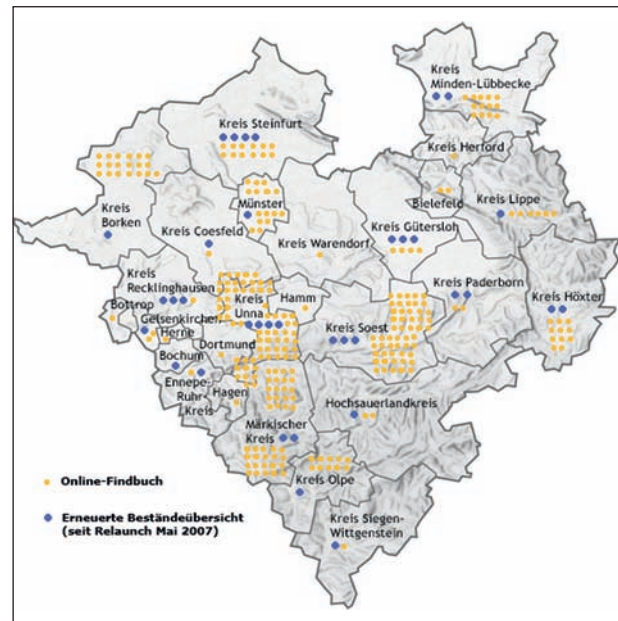


Abb. 2b: Westfälisch-Lippisches Engagement der Kommunalarchive bei „Archive in NRW“ (Stand Februar 2012)

Unklare Rechtssituation

Ohne Mark Steinerts Beitrag in diesem Heft vorgreifen zu wollen, gilt prinzipiell, dass im Bezug auf die Zugänglichkeit von Findbüchern die gleichen Maßstäbe anzulegen sind wie an das zugehörige Archivgut.³ Unterliegt dieses keinen Schutzfristen mehr, kann auch das Findbuch online gestellt werden. Mit Nachlassgebern und Deponenten sollte der Punkt „Online-Findbücher“ zu den Fragen gehören, die man bei der Übernahme regelt. Wenn einzelne Verzeichnungseinheiten schützenswerte Angaben enthalten, kann man vor dem Einstellen der Online-Findbücher die Situation durch eine Anonymisierung entschärfen. Manchmal empfiehlt es sich auch, bestimmte Klassifikationsäste, die zum Beispiel personenbezogene Einzelfallakten enthalten, ganz aus dem Online-Findbuch heraus zu nehmen.

Unzureichende Findbuchqualität

Die Qualität der Erschließung kann ein weiterer Hinderungsgrund sein, der eine Veröffentlichung im Internet unmöglich erscheinen lässt. Natürlich bedarf es gewisser Mindeststandards, wenn man die Erschließungsdaten online präsentiert. Es ist aber durchaus legitim, Teilfindbücher mit ersten Erschließungsergebnissen oder auch nur vorläufig klassifiziertes Material zu präsentieren und im Vorwort diesen Erschließungsstand dem Leser mitzuteilen. Gerade abgetippte ältere Findbücher aus dem vorletzten Jahrhundert mögen nicht unseren heutigen Ansprüchen genügen, doch auf eine Neuverzeichnung dieser Bestände zu warten, ist wohl auch keine Option. Ein Online-Findbuch bringt für den Nutzer auf jeden Fall bessere Recherchemöglichkeiten als in der Papierwelt, selbst wenn es nicht perfekt ist. „Mut zur Lücke!“ könnte man hier raten.

Fehlende Softwareunterstützung

Auch wenn die allermeisten Archive inzwischen EDV-gestützt erschließen, besitzen immer noch nicht alle eine ausgesprochene Archivsoftware mit geeigneten Exportfunktionen, um das Archivportal NRW zu beliefern. Hier greift ihnen das LWL-Archivamt gern unter die Arme: Sie liefern z. B. eine EXCEL-Liste mit Ihren Erschließungsdaten und wir machen ein XML-kodiertes Findbuch daraus und stellen es online. Auch im letzten Fall, dass nämlich keine Findbücher in Datenbankform vorliegen, kann man sich mit Bürosoftware helfen. Wir haben für den Bereich der Adelsarchive gute Erfahrungen mit Ehrenamtlichen gemacht, die eingescannte Findbücher im PDF-Format in EXCEL-Tabellen übertragen. Für eine einfache Aktenerschließung sind beispielsweise die Spalten „Klassifikation“, „Signatur“, „Titel“, „Enthält“, „Darin“, „Laufzeit von“, „Laufzeit bis“ und „Laufzeittext“ angelegt, jede Zeile umfasst die Daten einer Verzeichnungseinheit. Die erfassten Einträge werden dann zunächst durch den Archivar/die Archivarin gesichtet und anschließend in die Erschließungssoftware importiert. Auch bei Findbüchern auf Durchschlagpapier oder mit zahlreichen handschriftlichen Ergänzungen kommt man auf diese Weise zu guten Ergebnissen. Dieser Verfahrensweg hat für die Freiwilligen den Vorteil, dass sie die Unterstützung vom heimischen PC aus leisten können ohne ins Archiv kommen zu müssen. Für einen dieser Mitarbeiter ist es eine Beschäftigung bei seinen regelmäßigen Zugfahrten als Pendler, die er am Laptop erledigen kann. So kön-

³ Sehr eingehend hat die ARK-Arbeitsgruppe Archive und Recht das Thema „Bereitstellung elektronischer Findmittel in öffentlich zugänglichen Netzen“ bereits 2007 abgehandelt. Eine autorisierte Online-Fassung findet man auf den Seiten des Bundesarchivs unter http://www.bundesarchiv.de/imperia/md/content/bundesarchiv_de/fachinformation/ark/20070320_veroeffentlichungsgrundsaeetze_ark.pdf (Stand: 02.07.2012).

nen neben Freiwilligen auch Schreibkräfte der Verwaltung an der Retrokonversion mitarbeiten, ohne sich in die Programmumgebung eines Verzeichnungsprogramms einarbeiten zu müssen. Gewisse Arbeitsbelastungen kommen bei dieser Art der Retrokonversion auf das Archiv zu: Zum einen gibt es immer mal wieder Nachfragen, die zeitnah beantwortet werden müssen, zum anderen ist auch der genannte Korrekturlauf durch den Archivar/die Archivarin unvermeidlich. Bei sehr guten maschinenschriftlichen Vorlagen können Texterkennungsprogramme (sog. OCR-Software) erfolgreich eingesetzt werden. Jedoch erfordert die Aufteilung der Texte in die Datenbankfelder, die Fehlerkorrektur und auch die eine oder andere zu überarbeitende inhaltliche Information stets noch viel händische Arbeit, die auch recht monoton und anstrengend für die Augen sein kann. Welches Vorgehen den Erfassern mehr liegt, lässt sich bei Testläufen am Anfang des Projekts rasch ermitteln. Für die Motivation ist es wichtig, dass rasch greifbare Erfolge für die Beteiligten erkennbar sind: Ein abgetipptes Findbuch muss möglichst innerhalb von ein bis zwei Monaten online bereit stehen und eine Erwähnung der beteiligten Person im Vorwort sollte selbstverständlich sein.

Bevor die neuen Funktionen des Archivportals vorgestellt werden, soll auf eine weitere Möglichkeit hingewiesen werden, die es nach wie vor gibt: Das Retrokonversionsprojekt der DFG, das die Überführung von analogen Findbüchern in Internet-fähige Formate unterstützt.⁴ Auch über diesen Weg haben schon westfälische Archive ihre Findmittel für das Portal „Archive in NRW“ aufbereiten lassen. Auf eine Vielzahl von Praxisberichten aus staatlichen und kommunalen Archiven verweist die Internetseite der Koordinierungsstelle Retrokonversion.⁵

Das Portal „Archive in NRW“ – Hintergründe und Motive des Relaunchs

Das Portal „Archive in NRW“ ist das größte und eines der meistfrequentierten Archivportale in Deutschland. Über 490 Archive beteiligen sich zurzeit daran, fast 40 Mio. Anfragen registriert der Server pro Jahr, mehr als 100.000 pro Tag. Diese Zahlen sollen nicht (jedenfalls nicht nur) dazu dienen, um Eindruck zu schinden; auch nicht (oder nicht nur), um damit zu signalisieren, dass sich eine Beteiligung am Portal und eine Pflege der Archivinformationen dort lohnen. Die Zahlen machen vielmehr vor allem auf einen Trend aufmerksam: Von den jährlich 40 Mio. Anfragen auf dem Server richten sich über 13 Mio. auf die Beständeübersichten und Findmittel der Archive. Nun könnte man sagen: Das ist ja gerade einmal ein Drittel der Gesamtanfragezahl. Das wäre aber doch zu kurz gegriffen. Die Zahl der Zugriffe auf das Archivportal NRW ist in den letzten Jahren auf hohem Niveau einigermaßen konstant geblieben; mit dem Relaunch sind zu Beginn des Jahres die Anfragezahlen noch einmal gestiegen. Stetig gestiegen sind schon über die ganzen letzten Jahre die Zugriffszahlen auf Beständeübersichten und Findmittel, nämlich von 5 Mio. im Jahr 2008 auf heute eben 13 Mio. pro Jahr. Die Bestän-

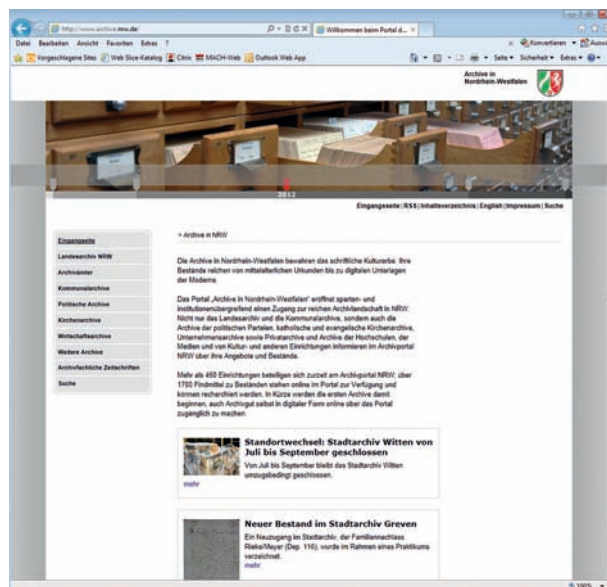


Abb. 3: Startseite des Internetportals „Archive in NRW“ nach dem Relaunch

dedatenbank ist mithin der Boom-Sektor des Portals. Und er boomt vor allem dort, wo umfangreiche und differenzierte Erschließungsinformationen im Portal vorliegen. Dies zeigt sich ganz deutlich an den Zahlen des Landesarchivs NRW, wo schon seit 2008 die Beständeinformationen weit- aus häufiger angefragt werden als die Seiten mit den allgemeinen Informationen. Auch bei den Portalseiten des Historischen Archivs der Stadt Köln zielen 90 % der monatlich etwa 35.000 Anfragen auf die über 700 eingestellten Findbücher. Damit dürfte deutlich werden: Die Hauptnachfrage der Portalbenutzer gilt den Beständeübersichten und Findbüchern, insbesondere über die spartenübergreifende Suche. Beständeübersichten und Findbücher sind der Kern und der Hauptsinn und Zweck einer archivischen Präsenz im Portal. Umso wichtiger ist es für die Archive, ihre Erschließungsinformationen ins Portal hineinzubekommen.

Mit dem Relaunch des Portals „Archive in NRW“ wurden die Import-Möglichkeiten für Beständeübersichten und Findmittel noch einmal verbessert und erweitert. Natürlich diene der Relaunch zunächst auch dazu, die Menüstruktur zu straffen, das Layout zu vereinfachen und so die Darstellung übersichtlicher und anwenderfreundlicher zu gestalten. Es sollten Funktionalitäten umgesetzt werden, die heute zum Standard von Internetpräsenzen gehören, und auch erste Schritte in Richtung auf das Web 2.0 unternommen werden: mit der Möglichkeit, Seiteninhalte der Archive zweisprachig anzubieten, Bilder als Galerien anzulegen, Nachrichten auch als RSS-Feeds zu verbreiten und den Standort des Archivs mit einfachen Mitteln über eine integrierte Google-Map anzuzeigen. Wichtiger als diese

4 Ansprechpartner ist die „Koordinierungsstelle Retrokonversion“ an der Archivschule Marburg, vgl. <http://www.archivschule.de/forschung/retrokonversion/> (Stand: 02.07.2012).

5 Vgl. direkt unter <http://www.archivschule.de/forschung/retrokonversion/publikationen/> (Stand: 02.07.2012).

Überarbeitung und Erweiterung der allgemeinen Funktionen war aber der Ausbau archivfachlicher Funktionalitäten im Portal. Das betrifft zwei Bereiche: zum einen den Findbuchimport, zum anderen die Einbindung von Digitalisaten.

Die Einbindung von Findmitteln im Archivportal NRW

Bislang gab es für den Findbuch-Import zwei Möglichkeiten:

- Ein Findbuch existierte im Format SAFT-XML. In diesem Fall verlief der Import in der Regel einfach und fehlerfrei. Wenn es Probleme gab, konnten diese in der Regel schnell mit kleineren Bearbeitungen der Ausgangsdatei behoben werden. Mehrere Erschließungsprogramme besitzen inzwischen SAFT-Schnittstellen. Dies gilt u. a. für die Produkte aus dem Hause Startext, aber auch für andere Produkte wie z. B. die Erschließungssoftware FAUST der Firma Land Software.
- Ein Findbuch existierte als eine aus dem Programm AUGIAS erzeugte Upload-Datei. Auch in diesem Fall konnte das Findbuch prinzipiell (ohne weitere Format-Konversion) ins Portal importiert werden – allerdings nicht immer einfach und nicht immer fehlerfrei. Wo Fehler auftauchten, konnten diese bislang nicht ohne Weiteres und auf die Schnelle, sondern oft nur durch unsere Fachleute beim Landesbetrieb Information und Technik NRW behoben werden.

Nach dem Relaunch stehen diese beiden Möglichkeiten des Findbuch-Imports nach wie vor zur Verfügung. Neu ist aber eine dritte Möglichkeit zum Import von Beständeübersichten und Findmitteln. Diese neue Möglichkeit trägt der wachsenden Bedeutung von EAD auch innerhalb der deutschen archivischen Fachgemeinschaft Rechnung. Dementsprechend bieten auch immer mehr Softwarehersteller EAD-Export-Funktionen in ihren Erschließungsprogrammen an. Das Internetportal „Archive in NRW“ kann ab sofort diese EAD-Dateien verarbeiten, im Prinzip wenigstens. Denn es fehlte bislang eine über die Länder und Sparten abgestimmte Konvention zur Verwendung des EAD-Standards, eine Verständigung auf ein einheitliches EAD-Profil. Inzwischen hat eine von der Archivreferentenkonferenz eingesetzte Bund-Länder-Arbeitsgruppe unter Beteiligung auch von Kommunalarchiven im Vorfeld der Planungen für das Archivportal-D im Rahmen der Deutschen Digitalen Bibliothek das Konzept für ein solches einheitliches EAD-Profil vorgelegt.⁶ Das Archivportal NRW kann grundsätzlich Daten nach diesem abgestimmten EAD-Profil verarbeiten, wenigstens in den Kernelementen. Darüber hinaus ist es aber in der Lage, auch EAD-Dateien aus dem Programm AUGIAS zu importieren. Vermutlich lassen sich auch EAD-Dateien aus anderen Quellen mit zufriedenstellendem Ergebnis ins das Portal übernehmen; hierzu bedarf es sicherlich noch einiger Testläufe und auch der Rückmeldungen durch die Kolleginnen und Kollegen der am Portal beteiligten Archive. Grundsätzlich ist allerdings davon auszugehen,

dass die neue EAD-Schnittstelle die Möglichkeiten zum Import von Findbüchern in das Portal erweitert und damit langfristig dazu beiträgt, mehr archivischen Inhalt für das Portal zu akquirieren.

Und wenn dereinst das Archivportal D im Rahmen der Deutschen Digitalen Bibliothek Wirklichkeit geworden sein wird? Dann wird das Archivportal NRW zugleich als Aggregator für dieses nationale Archivportal wirken können. Die technischen Voraussetzungen dafür sind schon jetzt weitgehend geklärt. Wenn die Archive dies wünschen, können Beständeübersichten und Findmittel – neue wie auch Bestandsdaten – automatisiert, so wie es aussieht, über eine Harvesting-Technologie an das Archivportal D weitergegeben werden. Die Archive selbst können dabei steuern, was weitergegeben werden soll und was nicht.

Einbindung von Digitalisaten im Archivportal NRW

Seit dem Relaunch besteht auch die Möglichkeit, Digitalisate von Archivgut in das Archivportal NRW einzubinden. Im Konzept des Portals war die Bereitstellung von Archivgut-digitalisaten eigentlich von Anfang an als dritte Ausbaustufe des Portals vorgesehen (nach den Beständeübersichten und Findmitteln); begrenzte Daten- und Leitungskapazitäten, zudem aber auch eine gewisse generelle fachliche Skepsis, ob angesichts der riesigen Mengen Archivgut jemals in nennenswertem Umfang digital genutzt werden kann, haben indes die Umsetzung verzögert. Inzwischen dürfte einigermaßen klar sein, dass langfristig hinter die Präsentation digitalen Archivguts im Netz kein Weg mehr zurückführt. Die an Google-Books und ähnliche Netzangebote gewöhnten Benutzer erwarten von den Archiven Digitalisate; umgekehrt können die Archive in Anbetracht bereits existierender Großprojekte zumindest unter technischen und organisatorischen Gesichtspunkten immer weniger triftige Gründe gegen eine Mengendigitalisierung ins Feld führen. Wie bei den Bibliotheken gibt es auch bei den Archiven einen massiven (im Landesarchiv NRW auch mit Daten zu untermauernden) Trend zur virtuellen Distanznutzung der Archive. Diesen Trend kann man begrüßen oder verteufeln, ändern kann man ihn wohl nicht. Das Landesarchiv NRW hat inzwischen etwa 0,8 % seiner Bestände digitalisiert; ein erheblicher Teil dieser Digitalisate steht den Nutzerinnen und Nutzern schon jetzt in den Lesesälen zur Verfügung. Die Zugriffszahlen auf diese Digitalisate haben sich mit dem wachsenden Angebot in den letzten drei Jahren mehr als verdoppelt. An dieser Stelle ist deshalb auch für das Internetportal mit einem neuen Boom zu rechnen, sobald die ersten Digitalisate online stehen.

Die technischen Voraussetzungen dafür sind jetzt jedenfalls geschaffen. Jede in einem Findbuch verzeichnete Archiveinheit kann im Portal mit einer beliebigen Zahl von Di-

⁶ Vgl. Ulrich Fischer/Sigrid Schieber/Wolfgang Krauth/Christina Wolf, Ein EAD-Profil für Deutschland. EAD(DDB) als Vorschlag für ein gemeinsames Austauschformat deutscher Archive, in: *Archivar* 65 (2012), S. 160–162.

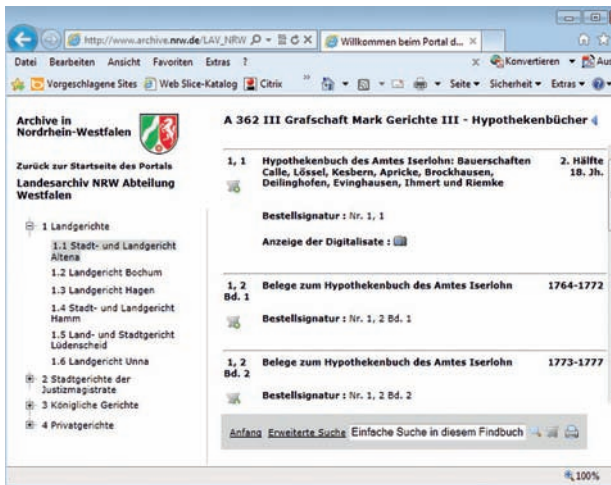


Abb. 4: Findbuchanzeige mit Verweis auf Digitalisate im Portal „Archive in NRW“

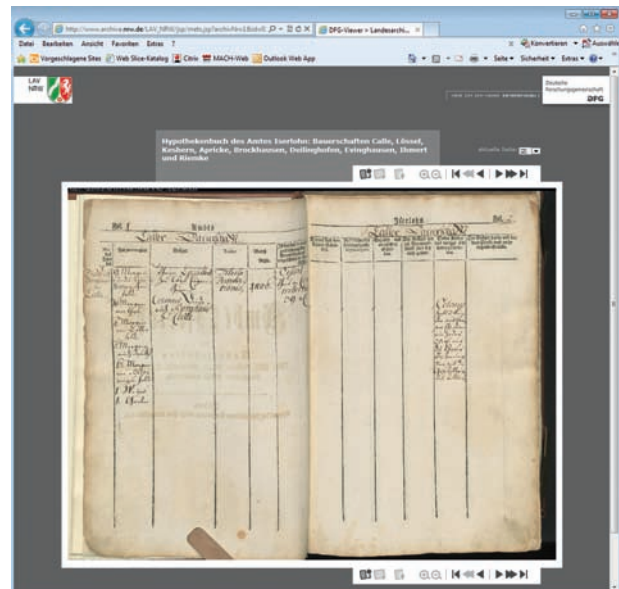


Abb. 5: Anzeige von Archivgutdigitalisaten im Portal „Archive in NRW“ über den zugeschalteten DFG-Viewer

gitalisaten verknüpft werden. Dazu müssen in einem ersten Schritt die entsprechenden Digitalisate auf einem öffentlich zugänglichen Webserver abgelegt werden. Alle Digitalisate, die zu einer Archiveinheit gehören, müssen in einer Datei im Format METS-XML referenziert werden. Zum Erzeugen solcher METS-Dateien gibt es frei verfügbare Tools; auch das Landesarchiv kann bei Bedarf gerne ein solches Tool kostenfrei zur Verfügung stellen. In einem zweiten Schritt müssen dann die Findbücher (sei in SAFT-XML oder EAD-XML) an vorgegebenen Stellen mit Links auf die METS-XML-Dateien versehen werden; solange die archivischen Software-Hersteller den METS-Standard für Digitalisatverknüpfungen nicht oder nur ansatzweise unterstützen, ist dieser Vorgang mit händischem Aufwand verbunden. Die Situation wird sich hoffentlich in Zukunft verbessern.

Bei der Anzeige der Findbücher holt sich das Archivportal automatisch die METS-Dateien und die Digitalisate vom Server ab und erstellt daraus mit Hilfe des DFG-Viewers eine entsprechende Anzeige, wobei Titelinformationen zur Archivguteinheit aus der Datenbank in die Darstellung übernommen werden. Wie beim Findbuch-Import ist die technische Infrastruktur so ausgelegt, dass die Archivgutdigitalisate später – ohne weitere Änderung – auch im Archivportal D angezeigt werden können – natürlich auch hier, nur, sofern die Archive dies wünschen.

Es ist ganz klar: Weder für die Findbuch- noch für die Digitalisateinbindung bietet das Archivportal NRW den teilnehmenden Archiven eine fertige Out-of-the-box-Lösung an. Der Wunsch, einfach das, was man vor Ort schon hat, unbearbeitet hochladen und weaternutzen zu können, ist zugegebenermaßen reizvoller; aber er ist unrealistisch. Verbundlösungen können nie eine Flexibilität für alle Formate besitzen; sie müssen immer Vorgaben machen. Was allerdings zu Recht verlangt werden kann, ist die Orientierung an etablierten Standards der Fachgemeinschaft. Dies

versucht das Archivportal NRW – jetzt vor allem durch die EAD-Schnittstelle, die Ausrichtung am METS-Datenprofil des DFG-Viewers und durch die Abstimmung mit dem Konzept für das Archivportal D. Als Portalbetreiber würden wir uns freuen, wenn zukünftig noch mehr Erschließungsinformationen und dann auch Archivgutdigitalisate über das Archivportal NRW den Weg ins Netz finden würden. Mit Unterstützung der Kolleginnen und Kollegen bei den Archivämtern werden sich sicher für viele Probleme, die bei der Online-Stellung von Archivinformationen entstehen, Lösungen finden. Wichtig ist aber, dass die Probleme offen benannt und kommuniziert werden, die momentan noch dazu führen, dass nur ein kleiner Teil der in den Archiven verfügbaren Informationen über das Portal online, für eine breite Öffentlichkeit zugänglich ist. ■



Dr. Andreas Pilger
Landesarchiv NRW, Fachbereich Grundsätze
andreas.pilger@lav.nrw.de



Dr. Peter Worm
LWL-Archivamt für Westfalen
peter.worm@lwl.org