

MASSENDIGITALISIERUNG

Eine Million Bücher online

IN EINER GROSS ANGELEGTEN PUBLIC-PRIVATE-PARTNERSHIP MIT GOOGLE WIRD DER GESAMTE URHEBERRECHTSFREIE BESTAND DER BAYERISCHEN STAATSBIBLIOTHEK DIGITALISIERT UND IM INTERNET ZUR FREIEN NUTZUNG BEREITGESTELLT.

VON KLAUS CEYNOWA

Im Frühjahr dieses Jahres haben die Bayerische Staatsbibliothek und der weltweit größte Internetsuchmaschinen-Betreiber Google einen auf mehrere Jahre angelegten Kooperationsvertrag unterzeichnet.

Im Rahmen dieser Vereinbarung wird Google den gesamten gemeinfreien Bestand der Bibliothek – das sind deutlich mehr als eine Million Bücher und Zeitschriftenbände – einscannen. Die Digitalisate werden über das Internet zur weltweiten Nutzung zur Verfügung gestellt, und zwar sowohl über die Website der Bayerischen Staatsbibliothek (www.bsb-muenchen.de) wie über die Buchsuche Googles (<http://books.google.com>) und die allgemeine Websuche Googles. Ausgenommen von dem Digitalisierungsprojekt sind nur die einzigartigen Handschriften- und Inkunabelbestände sowie seltene und besonders wertvolle historische Drucke.

Durch die Kooperation mit Google wird im Verlauf weniger Jahre ein großer Teil des europäischen schriftlichen Kulturerbes – hunderttausende literarische und wissenschaftliche Quellenwerke in vielen Sprachen sowie ein umfassender, systematisch aufgebauter Korpus an historischer Forschungsliteratur – für jeden Nutzer, der einen Internetanschluss besitzt, per Mausklick zugreifbar gemacht. Neben der Bayerischen Staatsbibliothek



liothek nehmen eine ganze Reihe renommierter Bibliotheken, unter anderem die Universitätsbibliotheken von Stanford, Michigan, Harvard und Princeton, die New York Public Library und die Bodleian Library der Universität Oxford, an Googles Bibliotheksprojekt teil.

Internationale Forschungsbibliothek im Netz

Welche Gründe haben die Bayerische Staatsbibliothek bewogen, diese in ihrer Größenordnung europaweit einzigartige Partnerschaft mit Google einzugehen? Die Bayerische Staatsbibliothek ist eine internationale Forschungsbibliothek, deren über 450 Jahre hinweg systematisch aufgebauter Bestand eine für Wissenschaft und Studium

einzigartige Ressource darstellt. Mit 95.000 Handschriften belegt die Bayerische Staatsbibliothek den vierten Platz unter allen Bibliotheken weltweit, im Bereich der Inkunabeln mit 20.000 Exemplaren sogar den ersten Platz. Ihr Bestand von 49.000 Zeitschriften, von denen viele auch elektronisch verfügbar sind, wird in Europa nur von der British Library übertroffen.

Aufgrund ihrer internationalen Ausrichtung hat die Bayerische Staatsbibliothek, im Unterschied etwa zu einer Universitätsbibliothek, keine lokal definierte „primäre Nutzergruppe“. Ihre Angebote richten sich vielmehr an Forscher, Studierende und Informationssuchende weltweit und werden auch entsprechend nachgefragt, wie die intensive Nutzung durch ausländische Wissen-

Die Bayerische Staatsbibliothek – ein Wissensspeicher des schriftlichen europäischen Kulturerbes.



Auch wertvollste Handschriften der Bayerischen Staatsbibliothek werden gescannt – aber nicht durch Google.

schaffler und die hohe Inanspruchnahme durch Fernleih- und Dokumentlieferdienste (500.000 bearbeitete Bestellungen in 2006) zeigt. Insofern ist das Internet für die Bayerische Staatsbibliothek gleichsam das natürliche Medium, erlaubt es doch den Zugriff auf Informationen für jedermann in aller Welt und rund um die Uhr. Es ist somit ihr vorrangiges strategisches Ziel, den einzigartigen historischen Bestand so rasch wie möglich zu digitalisieren und – ganz pragmatisch – für die Welt nutzbar zu machen, soweit dies die juristischen und technologischen Bedingungen zulassen.

Münchener Digitalisierungszentrum

Die Bayerische Staatsbibliothek verfolgt dieses Ziel seit 1997, dem Gründungsjahr des mit Fördermitteln der Deutschen Forschungsgemeinschaft aufgebauten Münchener Digitalisierungszentrums. Dieses Zentrum, das heute als

„Referat Digitale Bibliothek“ fest in die Arbeitsabläufe der Bibliothek integriert ist, hat seit seiner Gründung mehr als 80 zumeist drittmittelgeförderte Projekte durchgeführt, zum großen Teil kooperativ mit der universitären und außeruniversitären Forschung. Neben der Entwicklung innovativer Technologien und Workflows steht die Digitalisierung von Handschriften und Bavarica sowie von Beständen der DFG-geförderten Sondersammelgebiete, vor allem Geschichte und Osteuropa, im Vordergrund. Bis heute wurden mehr als 20.000 Titel mit insgesamt zehn Millionen Seiten gescannt und für die Internetpräsenz aufbereitet, was einem Datenbestand von über 40 Terabyte entspricht. Die „Digitalisierungsstraße“ der Bayerischen Staatsbibliothek verfügt gegenwärtig über zwölf hochauflösende Buchscanner, wozu auch zwei „Grazer Kameratische“ zur objektschonenden Digitalisierung wertvollster Handschriften und Drucke zählen. In einem aktuellen Projekt zur Digitalisierung der circa 40.000 deutschsprachigen Drucke des 16. Jahrhunderts im Bestand der Bayerischen Staatsbibliothek werden erstmalig zwei vollautomatisch arbeitende Scanner mit einem Leistungsvolumen bis zu 3.000 Seiten pro Stunde eingesetzt.

„Boutique“ Digitisation Projects versus Massendigitalisierung

Sieht man einmal vom zuletzt genannten Vorhaben ab, handelt es sich bei den Projekten des Münchener Digitalisierungszentrums – ähnlich verhält es sich mit den Aktivitäten anderer großer deutscher Bibliotheken – um thematisch oder materialspezifisch fokussierte Ansätze, die sich auf mengenmäßig relativ überschaubare Textkorpora richten, für die ein spezifisches Forschungsinter-

resse vorliegt oder die von ihrem Bestandsprofil her einzigartig sind. Ronald Milne von der British Library spricht hier von „Boutique“ Digitisation Projects. Mit der Kooperation der Bayerischen Staatsbibliothek mit Google wird dagegen im deutschen Bibliothekswesen erstmals ein Digitalisierungsprojekt in industriellem Maßstab, also eine echte Massendigitalisierung, technisch und logistisch in Angriff genommen. Eine „Auswahl“ der Bücher findet nur noch nach ihrer konservatorischen Eignung für den Scanprozess und hinsichtlich bestimmter Vorgaben nach Größe und Umfang statt, die durch Googles proprietäre Scantechnologie bedingt sind.

Ein auf mehr als eine Million Bücher berechnetes Massendigitalisierungsprojekt ist, wenn es im überschaubaren Zeitraum von etwas mehr als einem halben Jahrzehnt abgewickelt werden soll, finanziell nur im Rahmen einer Public-Private-Partnership zu bewältigen. Die aktuellen Bekanntmachungen und Empfehlungen der Europäischen Kommission und des Europäischen Rates zur „Digitalisierung und Online-Zugänglichkeit kulturellen Materials und dessen digitaler Bewahrung“ lassen keinen Zweifel daran, dass sich die Europäische Union nicht an den operativen Kosten der Massendigitalisierung von Kulturgut beteiligen wird, sondern dies als Aufgabe der Mitgliedstaaten betrachtet. Auch die derzeitigen Initiativen auf Bund-Länder-Ebene zur Entwicklung einer „Deutschen Digitalen Bibliothek“ als Beitrag zu einer „Europäischen Digitalen Bibliothek“ sind hinsichtlich der Frage, wer die finanzielle Last einer breit angelegten Digitalisierung des deutschen Kulturgutes tragen soll, noch völlig unbestimmt. Angesichts des immensen Finanzierungsbedarfs weist die Europäische Kommission explizit auf die Notwendigkeit von

Public-Private-Partnerships hin, um die Aufgabe der Digitalisierung der jeweiligen nationalen Kulturbestände zu bewältigen.

„Google Digital Copy“ und „Library Digital Copy“

Der entscheidende Vorteil der Kooperation mit Google liegt für die Bayerische Staatsbibliothek darin, dass Google die gesamten Digitalisierungskosten trägt. Google behält als Gegenleistung eine „Google Digital Copy“ zur Integration in die Google Buchsuche und die allgemeine Google Websuche. Die Bayerische Staatsbibliothek erhält ihrerseits die „Library Digital Copy“ zur Integration in ihre diversen Internetangebote. Dies ist der wesentliche Punkt im mit Google abgeschlossenen Vertrag, der hinsichtlich seiner Details einer bei Public-Private-Partnerships üblichen, im Unternehmenspersönlichkeitsrecht begründeten Verschwiegenheitspflicht unterliegt: Die Bayerische Staatsbibliothek besitzt, indem sie eine digitale Kopie der von Google erzeugten Daten erhält, diese Daten „physisch“ und damit dauerhaft. Sie kann die „Library Digital Copy“ uneingeschränkt im Rahmen ihres Online-Katalogs (OPAC) und ihrer Webangebote zur Verfügung stellen, beispielsweise in Form fachlich se-

lektierter, für eine wissenschaftliche Recherche speziell aufbereiteter Forschungskorpora.

Innerhalb der digitalisierten Bücher kann mittels komfortabler Navigationsinstrumente „virtuell“ geblättert werden, darüber hinaus wird die Mehrheit der digitalisierten Werke aufgrund fortschrittlicher optischer Zeichenerkennung im Volltext durchsuchbar sein. Da ausschließlich urheberrechtsfreier Bestand der Bayerischen Staatsbibliothek digitalisiert wird, steht dem Benutzer auch die Option zum Download oder zur Kopie des gesamten Werkes für wissenschaftliche Zwecke zur Verfügung. Vermittels der Metadaten der digitalisierten Bestände ist der Bayerischen Staatsbibliothek zudem die Einbindung der „Library Digital Copy“ in regionale, nationale und internationale Portale und Dienste uneingeschränkt möglich. Der von Google digitalisierte Bestand kann also, gegebenenfalls fachlich oder materialspezifisch gefiltert, in Services wie die „Bayerische Landesbibliothek Online“, das kulturwissenschaftliche Internetportal des Freistaates oder in die geplante „Europäische Digitale Bibliothek“ integriert werden. Hervorzuheben ist in diesem Zusammenhang die Nicht-Exklusivität des Vertrages

mit Google, die es der Bayerischen Staatsbibliothek offen hält, angesichts zukünftig fortschreitender Technologie, spezieller Nutzungsbedürfnisse oder gewandelter strategischer Rahmenbedingungen auch von bereits durch Google gescannten Titeln erneut Digitalisate zu erstellen.

Bestandserhaltung und Langzeitarchivierung

Als zentrale Landes- und Archivarbibliothek des Freistaates Bayern hat die Bayerische Staatsbibliothek für die langfristige Nutzbarkeit ihrer Bestände Sorge zu tragen. Die Bestandserhaltung gerade der vom Papierzerfall akut bedrohten Bestände aus dem 19. Jahrhundert kann durch die von Google geleistete Digitalisierung gleichsam „miterledigt“ werden, da durch Digitalisierung die Informationssicherung der zerfallsbedrohten Bücher gewährleistet wird. Aufgrund des fortgeschrittenen Schädigungsgrades ist hier der Zeitfaktor von besonderer Bedeutung. Auch für die nicht zerfallsbedrohten Bestände bedeutet die Digitalisierung einen besonderen konservatorischen Schutz, da im Regelfall künftig auf das Digitalisat zugegriffen wird, während das Original nur noch zu speziellen Forschungszwecken herangezogen werden muss. Die Langzeitsicherung der digitalisierten Bestände wird die Bayerische Staatsbibliothek in enger Abstimmung mit dem Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften vornehmen.

Scanvorgang einer empfindlichen Handschrift am „Grazer Kameratisch“ des Münchener Digitalisierungszentrums.



Vorteile für Google

Die Digitalisierung des gesamten urheberrechtsfreien Bestandes der Bayerischen Staatsbibliothek unternimmt Google als börsennotiertes Unternehmen selbstverständlich nicht aus altruistischen Motiven. Sieht man auf die Positionierung der Google Buchsuche im Ge-

samtportfolio Googles, so ist eine direkte Vermarktung der „Google Digital Copy“ offenkundig nicht vorgesehen. Sie erscheint auch wenig sinnvoll, wenn die Bibliothek die identische „Library Digital Copy“ über ihre Webangebote

das Kernprodukt Googles, nämlich die Internetsuche, „umrahmenden“ Dienste – Google Earth, Google Scholar, Google Mail und eben auch Google Booksearch – vor allem die Funktion zu haben, Google als das attraktivere Angebot gegenüber

Electronic Daily) im Oktober 2006 ausgeschrieben. Die dann abgeschlossene Kooperationsvereinbarung mit Google unterliegt der bei Verträgen öffentlicher Einrichtungen mit Wirtschaftsunternehmen gebräuchlichen Verschwiegenheits-



Die Magazine der Bayerischen Staatsbibliothek – 450 Jahre Sammeltätigkeit.

unentgeltlich zur Verfügung stellen kann. Der Zugriff auf die „Google Digital Copy“ erfolgt sowohl über die allgemeine Websuche Googles wie über den speziellen Webdienst der Google-Buchsuche. Zumindes für Letzteren ist keine die Rechercheergebnisse begleitende Online-Werbung vorgesehen.

Gleichwohl ist ein indirekter, aber dennoch signifikanter Wettbewerbsvorteil Googles durch das Bibliotheksprojekt gegeben. Durch die im Volltext erfassten Digitalisate wird der Suchindex Googles um immensen Mengen an Content bereichert, was im Endeffekt zu einer optimierten Recherche, mehr Nutzern, mehr Klicks auf den Anzeigen und damit höheren Erlösen führt. Insgesamt scheint die Vielzahl der

konkurrierenden Suchmaschinen erscheinen zu lassen, die immer nur „einen Mausklick entfernt“ sind. „Users like Google better“ – gemäß diesem Prinzip trägt auch die Buchsuche Googles zur Attraktivitätssteigerung der Suchmaschine für Onlinewerber bei.

Konservatorische Eignung und Qualitätsstandards

Die Bayerische Staatsbibliothek hat die geplante Massendigitalisierung ihres urheberrechtsfreien Bestandes als Dienstleistungskonzession in einem transparenten und diskriminierungsfreien Verfahren in Form einer „Bekanntmachung zur Teilnahme am Verhandlungsverfahren“ im Supplement des Amtsblatts der Europäischen Union (Tenders

pflicht, die Auskünfte zu technischen und organisatorischen Details des Vertrages ebenso ausschließt wie Informationen zum genauen Umfang der zur Digitalisierung vorgesehenen Bestände und zur exakten Projektlaufzeit. Dies hat zu kritischen Rückfragen vor allem hinsichtlich der Wahrung konservatorischer und qualitativer Anforderungen geführt, die an bibliothekarische Digitalisierungsprojekte üblicherweise gestellt werden.

Hier ist hervorzuheben, dass die konservatorischen Kriterien, die darüber entscheiden, welche Bücher von ihrem Zustand her zur Digitalisierung geeignet sind und welche nicht, grundsätzlich von der Bayerischen Staatsbibliothek gemeinsam mit Google festgelegt werden. In

Zweifelsfällen liegt die Entscheidung hierbei immer bei der Bibliothek. Aufgrund der konsequenten Einbeziehung des international renommierten „Instituts für Buch- und Handschriftenrestaurierung“ der Bayerischen Staatsbibliothek ist sichergestellt, dass kein Buch, das beim Scannen mit der von Google eingesetzten Technologie Schaden nehmen könnte, an Google übergeben wird. Auch die mit Google vereinbarten Qualitätsstandards bewegen sich in den für drittmittelgeförderte Digitalisierungsprojekte üblichen Toleranzbereichen. Hier ist es durchaus von Vorteil, dass die Bayerische Staatsbibliothek nicht zu den Erstteilnehmern am 2004 von Google initiierten Bibliotheksprojekt zählt, sondern als „Späteinsteiger“ von den kontinuierlich optimierten Qualitätskontrollen Googles profitieren kann.

Zukunft der Bibliotheken?

Ganz grundsätzlich wird mit Blick auf Googles Bibliotheksprojekt, an dem gegenwärtig bereits 26 große Bibliotheken mit mittlerweile insgesamt circa 30 Millionen zu scannenden Büchern teilnehmen, gefragt, ob dieses gewaltige Vorhaben nicht langfristig den Untergang der Bibliotheken bedeutet – zumal auch ein zunehmender Anteil der neu erscheinenden Literatur von den Verlagen in digitaler Form – als E-Journals und E-Books – angeboten wird. Werden die Nutzer noch die Lesesäle, Ausleihschalter und Sondersammlungen der Bibliotheken aufsuchen, wenn große Teile ihrer Bestände online bequem von jedem Internetarbeitsplatz aus abrufbar sind? Mit Blick auf die nicht selten unzumutbare Überfüllungssituation der Lesesäle der Bayerischen Staatsbibliothek – 2006 wurden mehr als eine Million Besucher gezählt – und begeistert angenommener Öffnungszeiten bis Mitternacht ist diese Sorge wohl auch langfristig unbegründet. Eher

kann man von einer Renaissance der Bibliotheken sprechen, die als Orte der wissenschaftlichen Kommunikation, des kulturellen Austausches, des konzentrierten Lernens und wissenschaftlichen Forschens sich einer kontinuierlich steigenden Nutzung erfreuen, obwohl zusehends größere Teile ihres Informationsangebotes online bereitgestellt werden.

Zuweilen wird auch bezweifelt, ob die der Bayerischen Staatsbibliothek zur freien Verfügung überlassene „Library Digital Copy“ überhaupt noch nachgefragt wird, wenn die identische „Google Digital Copy“ über die populärste Internet-Suchmaschine der Welt aufgerufen werden kann. Auch dieser Zweifel ist aus Sicht der Bayerischen Staatsbibliothek unbegründet. Die Bibliothek ist durchaus nicht unglücklich darüber, dass ihr urheberrechtsfreier Bestand künftig auch über Google und damit die weltweit meist genutzte Webrecherche zugänglich ist. Denn schließlich entspricht dies dem ureigensten Auftrag jeder Bibliothek: Menschen und Wissen miteinander in Verbindung zu bringen. Die Angebote der Google Buchsuche und der Bayerischen Staatsbibliothek sind eher als komplementär zu begreifen. Google geht es letztlich um die Anreicherung seines Suchindex mit Content, über den die Mitbewerber nicht verfügen und damit schließlich um die Sicherung seiner Marktführerschaft im Geschäft mit Online-Werbung. Die Aufbereitung großer digitaler Textkorpora für spezifische wissenschaftliche Nutzungsinteressen und ihre Einbettung in netzbasierte Forschungs- und Lernumgebungen, wie sie für das künftige Angebot der „Library Digital Copy“ durch die Bayerische Staatsbibliothek kennzeichnend sein werden, stellt zumindest nicht das Kerngeschäft Googles dar. Statt von einer Angebotskonkurrenz wird man also eher von einer Koexistenz unterschiedlich fokussier-



BSB

ter Dienste und Nutzungsinteressen ausgehen können.

Ab Anfang 2008 sollten die ersten von Google digitalisierten Titel der Bayerischen Staatsbibliothek im Netz stehen – mit dann sehr rasch wachsenden Quantitäten. Insgesamt versteht sich die Public-Private-Partnership der Bayerischen Staatsbibliothek mit Google als ein Beitrag zum Kernauftrag der Bibliothek: den einzigartigen historischen Bestand für zukünftige Generationen zu bewahren und zugleich den Zugriff für die gegenwärtige Generation zu optimieren.

Der Autor ist Stellvertreter des Generaldirektors der Bayerischen Staatsbibliothek.



Die Lesesäle der Bayerischen Staatsbibliothek – bestens besucht trotz Bestandsdigitalisierung.