

Zum Aufbau einer digitalen Dokumentensammlung im Ungarischen Institut München

Der Informationsserver „Quellen und Materialien zur Beziehungs- und Integrationsgeschichte in Ostmittel- und Südosteuropa im 20. Jahrhundert unter besonderer Berücksichtigung Ungarns, der Minderheitenfrage und der europäischen Integration“

Vor dem Hintergrund der deutschen und internationalen Bestrebungen zur Digitalisierung von Bibliotheksbeständen wurde mit Beginn des Jahres 1997 ein von der Deutschen Forschungsgemeinschaft (DFG) gefördertes Programm ins Leben gerufen. Sein Ziel ist, »die Nutzung der neuen Kommunikations- und Publikationstechniken zur Verbesserung der wissenschaftlichen Arbeitsbedingungen beim Zugriff und bei der Verarbeitung von Literatur sowie von wissenschaftlichen Daten und Informationen verstärkt zu fördern«. Dadurch soll gewährleistet werden, daß elektronische Texte direkt am Arbeitsplatz des Wissenschaftlers zur Verfügung stehen, so daß »in einem Kernbereich der Förderung wissenschaftliche Forschungsliteratur aus den Beständen von Bibliotheken digitalisiert und über Kommunikationsnetze zugänglich gemacht werden«. ¹ Mit diesem »Förderungsprogramm ›Retrospektive Digitalisierung von Bibliotheksbeständen‹ will die Deutsche Forschungsgemeinschaft (DFG) den Grundstock für eine ›Verteilte Digitale Forschungsbibliothek‹ legen«. ²

In der Zwischenzeit wurde eine beträchtliche Zahl von Projekten ³ in diesem Kontext genehmigt, die insgesamt eine sehr positive Resonanz, sowohl in der Wissenschaft als auch in den Öffentlichkeit, gefunden haben. Dies führte 1998 zur Gründung von zwei DFG-geförderten Service- und Kompetenzzentren für retrospektive Digitalisierung. Das eine ist an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen (Göttinger Digitalisierungszentrum, GDZ), ⁴ das andere an der Bayerischen Staatsbibliothek (Münchener Digitalisierungszentrum, MDZ) ⁵ angesiedelt. Diese Bestrebungen zur elektronischen Erfassung von Altbeständen sind

¹ Bericht der Arbeitsgruppe Technik zur Vorbereitung des Programms „Retrospektive Digitalisierung von Bibliotheksbeständen“ im Förderbereich „Verteilte Digitale Forschungsbibliothek“ [im folgenden: Bericht]. DFG. http://www.sub.uni-goettingen.de/ebene_2/vdf/endas.pdf.

² Praktische Hinweise zur retrospektiven Digitalisierung von Bibliotheksbeständen. DFG. http://www.dfg.de/foerder/formulare/1_521.htm. 26. Oktober 2000.

³ Eine Liste der von der DFG genehmigten Projekte mit Stand 14. Juni 2001 unter <http://gdz.sub.uni-goettingen.de/de/vdf-d/>.

⁴ Siehe unter <http://gdz.sub.uni-goettingen.de/index.shtml>.

⁵ Siehe unter <http://www.bsb-muenchen.de/mdz/index.htm>.

in internationale Bemühungen zur computergestützten Bereitstellung bibliothekarischer Bestände eingebunden.

Im Verlauf der Jahre 1999 und 2000 wurden auch an einigen Münchener Instituten mit ostwissenschaftlichen Bezügen – darunter auch am Ungarischen Institut München (UIM) – Überlegungen angestellt, inwieweit eine einheitliche digitale Plattform zur Kunde Ost- und Südosteuropas aufgebaut und betrieben werden kann. Anfang 2000 kumulierten diese Überlegungen in dem Plan, einen Server für die Geschichte Ost- und Südosteuropas zu entwickeln und diesen im Internet der Forschung und Öffentlichkeit zur Verfügung zu stellen. Laut dem Konzept-Entwurf vom Februar 2000 sollten insbesondere die Bereiche »Fachinformation und Fachdienste«, »Fachbibliographien« sowie »Erschließung von Quellen« mit entsprechenden Teilbereichen erarbeitet werden.

In den folgenden Monaten übernahmen das Osteuropa-Institut München, die Abteilung für Geschichte Ost- und Südosteuropas der Ludwig-Maximilians-Universität, die Bayerische Staatsbibliothek und das Herder-Institut in Marburg die Planungen und Beantragung des Servers. Der „Osteuropaserver: Virtuelle Fachbibliothek zur Geschichte und Zeitgeschichte Osteuropas“ (im folgenden: „Osteuropaserver“) wurde im Juni 2001 bei der DFG von den vier genannten Institutionen beantragt und um weitere Projekte – so auch um das hier beschriebene – ergänzt. Das vorliegende Projekt gilt als eigener Antrag und als gesondertes Digitalisierungsvorhaben entsprechend dem Konzept der »Verteilten Digitalen Forschungsbibliothek«, besitzt jedoch Schnittstellen und enge Verbindungen zum Gesamtprojekt des „Osteuropaservers“.

Parallel zu diesen Entwicklungen und Vorhaben beschloß das Bayerische Staatsministerium für Wissenschaft, Forschung und Kunst, einen geistes- und sozialwissenschaftlichen Forschungsverbund zu gründen. Dieser Forschungsverbund Osteuropa (ForOst)⁶ reiht sich in die bayerische Forschungslandschaft⁷ als einer von zwei geistes- und sozialwissenschaftlichen Forschungsverbänden⁸ ein und soll »neue Erkenntnisse über Wandel und Kontinuität in den Ländern Ost- und Südosteuropas [...] und über die Auswirkungen des Umbruchs in den Staaten des ehemaligen Ostblocks

⁶ Siehe hierzu <http://www.abayfor.de/forost>.

⁷ Zur Arbeitsgemeinschaft der Bayerischen Forschungsverbände (abayfor) siehe <http://www.abayfor.de>. »Das Konzept der Forschungsverbände steht seit 1988 für interdisziplinäre Projekte aus speziellen Bereichen anwendungsnaher Forschung in Bayern. In einem Verbund arbeiten die Wissenschaftler aus mehreren Universitäten und die Wirtschaft eng miteinander zusammen. Die Effizienz von Spitzenforschung wird so durch Bündelung des Forschungspotentials über Universitätsgrenzen hinaus gesteigert. [...] Die Verbände profitieren von den Synergien, die ein fachübergreifendes Kompetenznetzwerk mit sich bringt.« <http://www.abayfor.de/frames/ueberuns.html>.

⁸ Der andere geistes- und sozialwissenschaftliche Verbund ist der Bayerische Forschungsverbund Area-Studies (FORAREA): <http://www.forarea.de>.

insbesondere auf Bayern [...] liefern«. ⁹ Sein Ziel ist, »Grundlagen für konkrete wirtschafts-, sozial- und kulturpolitische Empfehlungen zu schaffen«. ¹⁰ Der für eine zunächst zweijährige Laufzeit konzipierte Verbund wurde mit 2,6 Millionen DM ausgestattet, umfaßt 18 Teilprojekte und ist in seiner Ausgestaltung bisher einmalig in der Bundesrepublik Deutschland.

Im Rahmen dieser Bestrebungen – dem Aufbau eines Osteuropaservers sowie der Gründung von ForOst – wurde am UIM ein Projekt entwickelt, das sich in beide Aufgabenbereiche nahtlos integriert. Im folgenden werden dieses als Teilprojekt von ForOst genehmigte und im März 2001 unter Leitung des Verfassers angelaufene Projekt skizziert und der Stand der Forschungsarbeiten sowie die Perspektiven dargestellt.

Thematische Überlegungen

Parallel zu den bereits erwähnten Überlegungen zum Aufbau eines fachbezogenen Internetservers begann am UIM die Konzipierung eines ähnlich gearteten Projekts, das den veränderten Anforderungen an wissenschaftliches Arbeiten mittels moderner Kommunikationstechnologien und zum Zweck der beratenden Dienstleistung für Politik, Wirtschaft und Öffentlichkeit entsprechen soll. Denn der politisch-ökonomische Wandel in Ostmittel- und Südosteuropa seit Beginn der 1990er Jahre ließ nicht nur das wissenschaftliche Interesse an diesen europäischen Regionen wachsen, sondern auch dasjenige der breiten Öffentlichkeit. Die politisch-gesellschaftlichen Probleme und Konflikte, die sich im Zuge des Zusammenbruchs des ehemaligen Ostblocks entwickelten, betrafen zu einem großen Teil auch den Westen Europas und wirken bis in die Gegenwart nach. Auch die Schwierigkeiten der wirtschaftlichen Transformation sowie die unsichere zukünftige Entwicklung einerseits, und das gewachsene Interesse an den betreffenden Regionen andererseits ließen die Zahl der wissenschaftlichen Publikationen nicht nur in den Ländern Ost- und Südosteuropas, sondern auch in den westeuropäischen und nordamerikanischen Staaten drastisch ansteigen. Dies führte zu einer nahezu unüberschaubaren und kaum recherchierbaren Flut von Veröffentlichungen über Ost- und Südosteuropa, die zusammen mit den bereits bestehenden Quellen und Materialien einen bedeutenden Anteil der historischen Forschung ausmachen.

Auch wenn für Politik, Wirtschaft und Öffentlichkeit vor allem aktuelle Fragestellungen im Vordergrund stehen, spielt die Geschichtswissenschaft gerade für diese Regionen eine besondere Rolle, da hier in kaum zu über-

⁹ Bayerisches Staatsministerium für Wissenschaft, Forschung und Kunst, München. Pressemitteilung 28/2001, 28. Februar 2001.

¹⁰ Ebenda.

schätzendem Maße historische Vorgänge tagespolitische Entscheidungen beeinflussen und bestimmen. Als Beispiel hierfür kann die serbische Haltung in der Kosovo-Frage angeführt werden, die mit dem Mythos von der Schlacht auf dem Amselfeld (*Kosovo polje*) Tagespolitik betrieb und durch eine »nationale Erinnerungskultur« ethnische Spannungen hervorrief.¹¹ Diese Politiker »prägen, polen und operationalisieren ›Geschichte‹ also in ihrem Sinne.«¹² Geschichte spielt aber auch in anderen Staaten des ostmittel- und südosteuropäischen Raumes eine die Aktualpolitik zumindest mitbestimmende Rolle. So wirkt das Trianon-Trauma nicht nur in Ungarn nach, sondern auch in einigen Nachbarstaaten,¹³ die Ungarn trotz der mehrfachen Anerkennung bestehender Grenzen durch Budapest mißtrauen und revisionistische Forderungen befürchten. Aber auch andere historische Ereignisse, die sich in nationalen Mythen widerspiegeln, sei es in dem Traum von einem Groß-Serbien, Groß-Griechenland, Groß-Albanien, Groß-Bulgarien oder einem Vereinigten Makedonien finden sich in heutigen politischen Handlungen oder Plänen wieder. »Mythisches Denken [...] impliziert immer auch die aktionelle Umsetzung der Mytheme.«¹⁴ Zwischenstaatliche sowie interethnische Probleme und Konflikte unterschiedlicher Ausprägung und Intensität – in der Hauptsache zwischen Staaten, deren Staatsgrenzen sowie ethnischen Trennlinien im 20. Jahrhundert geändert wurden – beinhalten ein Spannungspotential, das die gesamteuropäische Stabilität nachteilig beeinflussen kann und eine weitergehende europäische Integration erschwert, wenn nicht sogar verhindert.

In der gegenwärtigen Politik der ostmittel- und südosteuropäischen Staaten spielt also das historische Argument in der Tagespolitik eine viel größere Rolle als im Westen Europas. Vor diesem Hintergrund kommt der historischen Ost-, Ostmittel- und Südosteuropaforschung eine große Bedeutung zu, da sie die entwicklungsgeschichtlichen Grundlagen für gegenwärtige Fragestellungen liefern und somit zur Lösung aktueller Probleme einen bedeutenden Anteil beisteuern kann. Darüber hinaus kann sie das Bewußtsein westeuropäischer Entscheidungsträger aus Politik und Wirtschaft schärfen und so auf ein besseres Verständnis dieser Einflußfaktoren, denen ostmittel- und südosteuropäische Politiker in nicht zu unterschätzendem Maße unterworfen sind, hinwirken.

Doch nicht nur die Differenzen zwischen den Staaten und Völkern Ost-, Ostmittel- und Südosteuropas stellen ein Integrationshindernis dar, sondern auch die Kommunikationsprobleme einerseits zwischen den Politikern und Nationen Westeuropas, andererseits zwischen ihren Gegenparts in den betreffenden Regionen. Die seit dem Beginn der Umwälzungen in

¹¹ Vgl. Stefan Troebst: Politische Entwicklung in der Neuzeit. In: *Südosteuropa*. Ein Handbuch. Hgg. Magarditsch Hatschikjan, Stefan Troebst. München 1999, 73-102, hier 75.

¹² Ebenda.

¹³ Insbesondere in Rumänien und der Slowakei.

¹⁴ Reinhard Lauer: Literaturen. In: *Südosteuropa* 417-436, hier 426.

diesem Raum vergangenen Jahre haben verdeutlicht, daß die politische Verständigung zwischen west- und osteuropäischen Entscheidungsträgern hauptsächlich an einem beiderseitigen Informationsmangel sowie an unterschiedlichen politischen Kulturen und verschiedenen Auffassungen von der tagespolitischen Wichtigkeit historischer Gegebenheiten und Rezeptionen scheitert. Vor diesem Hintergrund kommt der historischen Ost-, Ostmittel- und Südosteuropaforschung nicht nur eine Informationsbereitstellende Rolle, sondern vor allem eine herausragende Vermittlerrolle zu. Dies geschah in der Vergangenheit hauptsächlich durch herkömmliche Publikationen und wissenschaftliche Veranstaltungen. Jedoch ist die Reichweite und vor allem die schnelle Verfügbarkeit derartiger Informationsdienstleistungen begrenzt. Ein frei zugängliches, wissenschaftlich fundiertes und redaktionell geprüftes Fachinformationssystem, das zudem in die Neuen Medien¹⁵ integriert ist und diese als Kommunikations- und Informationskanal nutzt, kann einen ergänzenden und kaum zu überschätzenden Beitrag zu Annäherung und gegenseitigem Verständnis in Europa liefern.

Eine sehr wichtige und verlässliche Grundlage zur Analyse und Integration des historischen Elements in einen westlichen Deutungs- und Erklärungszusammenhang bieten die Quellen und Materialien zur Beziehungs- und Integrationsgeschichte Ostmittel- und Südosteuropas im 20. Jahrhundert.¹⁶ Das UIM, dessen allgemeines Schwerpunktthema „Ungarn in den Beziehungsgeschichten Ost-, Ostmittel- und Südosteuropas“¹⁷ ist, beschloß angesichts der Möglichkeiten, die beide Forschungseinrichtungen – DFG und ForOst – boten, ein Forschungsprojekt aus dem Bereich der Digitalisierung auf den Weg zu bringen, das auf die Erfassung maßgeblicher Quellen aus diesem Bereich der neuesten Geschichte ausgerichtet ist. Der daraufhin konzipierte „Informationsserver ‚Quellen und Materialien zur Beziehungs- und Integrationsgeschichte in Ostmittel- und Südosteuropa im 20. Jahrhundert‘“ wurde im April 2000 beim Bayerischen Staatsministerium für Wissenschaft, Forschung und Kunst eingereicht. Die im Dezember 2000 tagende Gutachterkommission, die über die Projektanträge zur Gründung von ForOst zu entscheiden hatte, genehmigte den Projektantrag und empfahl zur Erhöhung der Synergieeffekte innerhalb von ForOst eine thematische Fokussierung des Vorhabens auf ein anderes Teilprojekt

¹⁵ In der Hauptsache sind das digitale Veröffentlichungen auf CD-ROMs und Publikationen im Internet.

¹⁶ Selbstverständlich sind auch ältere Quellen von größter Wichtigkeit, doch kann in diesem Zusammenhang darauf nicht eingegangen werden. Eine zeitliche und thematische Einschränkung des Forschungsbereiches, der hier skizziert wird, war im Interesse der Umsetzbarkeit unbedingt notwendig.

¹⁷ Zsolt K. Lengyel: Die Evaluierung des Ungarischen Instituts München e. V. (UIM). Die erste Phase der Konsolidierung und des Ausbaus. In: Ungarn-Jahrbuch 24 (1998/1999) 445-448, hier 446. Die Skizze des thematischen Institutsprofils vom gleichen Autor unter <http://ungarisches-institut.de>.

im Forschungsverbund. Es sollte eine enge Kooperation mit dem Vorhaben „Datenbank zur Minderheitenproblematik und zu den ethnischen Gruppen Südosteuropas (Ethnoloc Südosteuropa)“¹⁸ unter dem Projektleiter Gerhard Seewann (Südost-Institut München) und eine stärkere thematische und konzeptionelle Fokussierung hierauf erfolgen. Infolge der Umsetzung dieser Empfehlung, die Bestandteil der ministeriellen Bewilligung war, erhielt das UIM-Projekt den Titel „Quellen und Materialien zur Beziehungs- und Integrationsgeschichte in Ostmittel- und Südosteuropa im 20. Jahrhundert unter besonderer Berücksichtigung Ungarns, der Minderheitenfrage und der europäischen Integration“ (im folgenden: „Informationsserver“).

Forschungsziele

Ziel des nunmehr thematisch eingegrenzten Projekts ist die digitale Volltextfassung und (text-) strukturierte Aufbereitung von Quellen und Materialien, denen eine zentrale Bedeutung für die Beziehungs- und Integrationsgeschichte in Ostmittel- und Südosteuropa im 20. Jahrhundert zukommt. Diese sind vor allem Verträge, Abkommen, diplomatische Schriftstücke, Memoiren, Ansprachen oder Vorträge politischer Entscheidungsträger, aber auch Statistiken zu Wirtschaft, Gesellschaft sowie den nationalen Minderheiten. Besonderes Augenmerk wird auf diejenigen Quellen und Materialien gelegt, die Ungarn unmittelbar oder mittelbar betreffen, die Frage der europäischen Integration Ostmittel- und Südosteuropas behandeln und sich mit der Minderheitenproblematik dieses Raumes befassen. Gleichzeitig sollen Quellen eine besondere Beachtung finden, die zum Verständnis gegenwärtiger politischer und sozialer Probleme und Konflikte beitragen und die Wandel und Kontinuität – in zeitlicher und räumlicher Dimension – besonders deutlich zum Ausdruck bringen. Aufschlußreiche Quellen in diesem Sinne bilden beispielsweise die Friedensverträge nach den beiden Weltkriegen, internationale und bilaterale Minderheitenschutzverträge, innerstaatliche Minderheitenschutzbestimmungen, Wirtschafts- und Handelsabkommen mit außen- und/oder minderheitenpolitischen Implikationen, Verträge der Europäischen Gemeinschaft beziehungsweise Europäischen Union mit den Transformationsstaaten der untersuchten Region, Protokolle über Grenzveränderungen, national- oder minderheitenpolitisch aussagekräftige Denkschriften und Memoiren sowie Bevölkerungs- und Wirtschaftsstatistiken.

Insbesondere im Bereich der Minderheitenthematik ergeben sich seit Projektbeginn hohe Synergieeffekte durch eine enge Kooperation mit dem

¹⁸ Kurzbeschreibung und relevante Informationen unter <http://www.abayfor.de/cgi-bin/1/webcon/abayfor/details/31/Projekte/num/617/1>.

Projekt „Ethnodoc Südosteuropa“. Diese manifestieren sich zum einen in der Vermeidung einer Doppelbearbeitung von Quellen und Materialien durch eine frühzeitige Absprache und Koordinierung der Arbeitsschritte. Zum anderen bringt die dauerhaft vernetzte Vorgehensweise in diesem thematischen Teilbereich eine wechselseitige Arbeitserleichterung und ermöglicht die Bearbeitung einer weitaus größeren Zahl von Quellen, als es ohne die enge Kooperation möglich wäre. Durch einen gegenseitigen Austausch der bearbeiteten Volltextdokumente steht beiden Projekten die größtmögliche Anzahl von Quellen zur Verfügung. Die koordinierende Funktion des beantragten „Osteuropaservers“ und der institutionalisierte Datenaustausch innerhalb dieses Gesamtprojekts legt für den Fall seiner Bewilligung auch die Grundlage für eine kontinuierliche und längerfristige Kooperation, die den zukünftigen Anwendernutzen gewährleistet.

Der „Informationsserver“ soll im Rahmen des „Osteuropaservers“ die Nachfrage nach aussagekräftigen und für Forschung und Öffentlichkeit herausragenden Quellen und Materialien befriedigen. Mit der Volltextfassung zahlreicher Dokumente soll der „Informationsserver“ als wichtiges Nachschlagemedium für den vorliegenden Themenbereich etabliert werden und durch ein vielfältiges, redaktionell geprüftes und beständiges Informationsangebot den Anforderungen an eine digitale Forschungsbibliothek entsprechen. Die Hauptunterschiede zu den meisten bestehenden digitalen Quellensammlungen im Internet bestehen darin, daß hier eine punktuelle Sammlung für eine bestimmte Forschungsfrage in einem definierten Zeitraum herausgegriffen wird und so durch eine größere Anzahl von Dokumenten ein umfassenderes Bild gezeichnet werden kann.

Das digital aufbereitete Informationsangebot geht durch die Volltextfassung der Texte und die Volltextrecherche weit über die Möglichkeiten eines herkömmlichen Bibliothekskatalogs und einer Quellenedition hinaus. Erstens ist die Erreichbarkeit der Quellen unabhängig von Zeit und Ort und lediglich durch geringe technische Anforderungen – notwendig sind nur ein moderner Computer und ein Internetzugang – beschränkt. Zweitens bietet das angestrebte System die Möglichkeit einer dokumentübergreifenden, wortbezogenen Volltextrecherche und beschleunigt so die Suche nach geeigneten Dokumenten. Drittens ermöglicht die wissenschaftliche Aufbereitung und Strukturierung der Materialien inhaltlich sowie fachlich spezifizierte Auswahlkriterien. Diese genügen allen Rechercheanforderungen und sind zugleich offen, so daß mögliche technische Weiterentwicklungen nachträglich implementiert werden können. Dies verspricht eine dauerhafte Verfügbarkeit des Dienstes, sofern die weitere Finanzierung gesichert wird. Neben den unterschiedlichen Recherche- und Erschließungsmöglichkeiten soll das Projekt auch die redaktionelle Qualität des „Informationsservers“ auf hohem Niveau etablieren, um dessen Zitierfähigkeit zu gewährleisten. Zudem muß das System den sprachlichen wie orthographisch-satztechnischen Anforderungen entsprechen, die

sich bei der Erforschung des Zielraumes ergeben. Die Einführung der korrekten Schreibweisen, die nicht nur bei Quellen und grundlegenden Materialien aus juristisch relevanten Bereichen eine Rolle spielen darf, ist ein Novum in der digitalen Volltextwiedergabe quellenartiger Dokumente.

Methode

Die zu bearbeitenden Quellen liegen in der Regel nur in gedruckter Form – beispielsweise in Gesetzblättern oder Periodika – vor, während digitalisierte Dokumente eher die Ausnahme sind. Das gedruckte Ausgangsmaterial muß zur Weiterverwendung im „Informationsserver“ zuerst digitalisiert werden. Bei dieser Computererfassung sind prinzipiell zwei Methoden beziehungsweise Stufen möglich, nämlich die Erfassung eines Dokuments als Bild sowie als Text, wobei letztere noch weiter unterteilt werden kann. Die erste Stufe der digitalen Konversion ist die reine Bilderfassung mittels eines Scanners¹⁹ – das manuelle Abschreiben der Texte kommt in diesem Fall aus zeitlichen, finanziellen und personellen Gründen nicht in Betracht. In dieser Bearbeitungsphase wird der gedruckte Text in ein in Pixel (Bildpunkte) zerlegtes Bild (Image) umgewandelt und stellt Dokumente in graphischer Form als Bilddatei²⁰ – dem sogenannten digitalen Master – oder in dem Austauschformat PDF²¹ dar.

In dieser Bearbeitungsstufe können die Texte bereits der Öffentlichkeit zugänglich gemacht werden, genügend aber nicht den Ansprüchen eines Informationsservers. Diese Formate können nicht oder nur unter größten Schwierigkeiten und mit einer hohen Fehlerquote durchsucht werden, so daß zu ihrer Verwaltung und Strukturierung zusätzlich eine Datenbank sowie deren redaktionelle und inhaltliche Betreuung notwendig wäre. Die Alternative einer Linkliste, die auf die entsprechenden Bilddateien verweist, ist angesichts der Vielzahl von Dokumenten ebenso umständlich in der Benutzung wie in der Weiterbearbeitung und Datenpflege. Aus diesen Gründen implementiert das Forschungsprojekt des UIM auch die zweite

¹⁹ »Der Scanner ist ein Lesegerät, das über eine geeignete Software (gedruckte) Vorlagen für die Weiterverarbeitung mit einem Computer in maschinenlesbare Form umwandelt. Er wird als Peripheriegerät an den Computer angeschlossen.« *Bericht 7.*

²⁰ Für bitonale Vorlagen wie reine Textquellen kommen Bildformate wie TIFF (*Tagged Image File Format*) mit einer verlustfreien Komprimierung der Gruppe 4 oder PNG (*Portable Network Graphics*), das sich zunehmend im World Wide Web etabliert. »Aus Sicht der Arbeitsgruppe kommen beide genannten Formate für Digitalisierungsvorhaben in Frage, wobei TIFF bei abgeschlossenen und derzeit laufenden Digitalisierungsvorhaben mit Abstand am häufigsten eingesetzt wird.« *Bericht 14.* Für den Onlinezugriff auf die Bilddaten kommt TIFF jedoch nicht in Frage, sondern es müssen PNG oder die Formate GIF (*Graphics Interchange Format*) beziehungsweise JPG/JPEG (*Joint Photographic Experts Group*) verwendet werden.

²¹ Das PDF-Format (*Portable Document Format*) ist ein plattformunabhängiges Format für den Dokumentenaustausch, in den das Bitmap-Image, also das Bild des Textes eingebunden wird.

Stufe der digitalen Konversion, nämlich die Volltexterfassung. Da das manuelle Erfassen aus den genannten Gründen nicht möglich ist, müssen die erfaßten Bilddateien mit einem Texterkennungsprogramm – einem sogenannten OCR-Programm (*Optical Character Recognition*) – weiterbearbeitet werden. Hierbei liest und analysiert das Programm die vom Scanner gelieferten Bildinformationen und sucht in ihnen wesentliche Merkmale – das sind Bögen, Kreise, Punkte – die es den Buchstabenzeichen zuordnen kann. Die im Projekt verwendete OCR-Software²² beinhaltet 121 Erkennungssprachen, so daß auch die für die vorliegende Fragestellung relevanten Dokumente korrekt erfaßt werden können. Die Erkennungsgenauigkeit der Software liegt bei einer guten Vorlage bei etwa 98 bis 99 Prozent, was jedoch bei einem Dokument mit 10.000 Zeichen immer noch eine Fehlerquote von 100 bis 200 Zeichen bedeutet. Bei älteren und schlechteren Vorlagen liegt die Fehlerquote deutlich höher. Um jedoch dem Anspruch der Zitierfähigkeit der Dokumente zu genügen, müssen die eingescannten und mittels OCR erkannten Dokumente mehrfach nachbearbeitet und mit dem Original abgeglichen werden.

Die derart bearbeiteten Dokumente müssen anschließend in einem Unicode-Format²³ abgespeichert werden, um die eventuell vorhandenen diakritischen Zeichen zu erhalten und die Weiterbearbeitung in einem Textverarbeitungsprogramm zu ermöglichen. Der Text liegt dann zwar in redaktionell bearbeiteter Form vor, die Suchfunktionen können jedoch noch nicht in der beabsichtigten Form implementiert werden. Die Erweiterung dieser Digitalisierungsstufe ist die Strukturbeschreibung, die auf den Volltext aufsetzt. »Unter Strukturbeschreibung von Dokumenten versteht man die formatunabhängige Kennzeichnung bzw. Markierung von distinktiven strukturellen Elementen eines Textes, wie Überschrift, Absatz etc. Beschrieben wird somit die logische Struktur eines Dokumentes, weniger sein Layout.«²⁴ Hierfür werden verschiedene Beschreibungssprachen eingesetzt, von denen die bekannteste das HTML (*Hypertext Markup Language*) ist. Für die hier gestellten Anforderungen ist HTML – das auf SGML (*Standard Generalized Markup Language*) aufbaut – jedoch nicht ausreichend, da es nur bedingt Strukturen wiedergibt, sich nur eingeschränkt für den Datenaustausch eignet und keinen robusten Link-Mechanismus hat.²⁵ Im Unterschied hierzu hat die Metasprache SGML ein wesentlich differenzierteres Beschreibungsinstrumentarium, eignet sich aber für das Projekt

²² ABBYY Finereader 5.0.

²³ Unicode ist ein vom Unicode-Konsortium entwickelter Standard, der die Zeichencodierung steuert und ein internationales, erweiterbares 16-Bit-Zeichencodierungssystem für die Informationsverarbeitung bietet, das die wichtigsten Sprachen der Welt umfaßt. Beim Speichern kommen in dem vorliegenden Fall die Formate RTF (*Rich Text Format*) und DOC (Word-Document) in Frage.

²⁴ Bericht 18.

²⁵ Siehe hierzu Simon North – Paul Hermans: XML in 21 Tagen. München 2000, 18-19.

ebenfalls nicht, da es zu teuer und zu kompliziert für den Einsatz im World Wide Web ist. So fiel die Wahl auf XML (*Extensible Markup Language*), eine Meta-Auszeichnungssprache zur Beschreibung strukturierter Daten, die auch ein Unterprofil von SGML ist. Ihr Vorteil besteht darin, daß diese Meta-Sprache auf verschiedenen Betriebssystemplattformen und in verschiedenen Anwendungen eingesetzt werden kann. »XML beschreibt eine Klasse von Datenobjekten, genannt XML-Dokumente, und beschreibt teilweise das Verhalten von Computer-Programmen, die solche Dokumente verarbeiten. [...] Ein Software-Modul, genannt XML-Prozessor, dient dazu, XML-Dokumente zu lesen und den Zugriff auf ihren Inhalt und ihre Struktur zu erlauben.«²⁶ Mit Hilfe dieses Zugriffs auf die Struktur und den Text der Dokumente werden die bearbeiteten Quellen des Informationsservers umfassend recherchierbar und erfüllen die Anforderungen an einen Informationsserver.

Mit der Strukturierung der Dokumente sind die Arbeiten im Rahmen des bei ForOst beantragten UIM-Projektes beendet. Zur Veröffentlichung der aufbereiteten Dokumente sind jedoch noch einige Schritte notwendig, die mit den bisher bewilligten Projektmitteln nicht vorgenommen werden können. Diese sind erst nach Bewilligung der Mittel durch die DFG sowohl für dieses Projekt als auch für das Oberprojekt „Osteuropaserver“ möglich. Die Publizierung der Dokumente ist nämlich abhängig von der Bereitstellung eines Internetserver beziehungsweise dem benötigten Speicherplatz auf einem Internetserver, auf dem die Daten abgelegt werden können. Darüber hinaus ist die Installation eines Suchprogramms sowie die Programmierung der entsprechenden Suchfunktionen notwendig, die in den Aufgabenbereich des „Osteuropaservers“ fallen. Somit ist die *Verfügbarkeit* und *Benutzbarkeit* des „Informationsservers“ von dem Gesamtvorhaben „Osteuropaserver“ abhängig.

Stand der Arbeiten und Ausblick

Mit den über ForOst bewilligten Mitteln konnte eine – vorläufig auf zwei Jahre befristete – halbe wissenschaftliche Mitarbeiterstelle am UIM eingerichtet werden. Darüber hinaus stehen – angesichts der Arbeitsziele – relativ geringe Mittel für wissenschaftliche Hilfskräfte sowie für technische Ausrüstung und laufende Sachausgaben zur Verfügung.

Der Projektbearbeiter (Andreas Schmidt-Schweizer) hat mit Beginn der Bearbeitungszeit im März 2001 eine Vorauswahl der zu digitalisierenden Quellen getroffen und eine umfangreiche Liste wichtiger Dokumente er-

²⁶ „XML 1.0 Recommendation“ des W3C (*World Wide Web Consortium*) vom 10. Februar 1998. Deutsche Übersetzung von Henning Behme und Stefan Mintert unter <http://www.mintert.com/xml/trans/REC-xml-19980210-de.html>.

stellt. Seither sind auch zwei wissenschaftliche Hilfskräfte für die Quellenbeschaffung aufgrund der eben erwähnten Liste sowie die ersten beiden Stufen der Digitalisierung verantwortlich. Sie scannen die Vorlagen ein und wandeln das Bild mittels OCR-Software in einen Text um. Dieser wird von den Hilfskräften anhand des Originals auf Übereinstimmung verglichen, um die redaktionelle Zuverlässigkeit zu gewährleisten, wobei insbesondere auf die korrekte Wiedergabe der diakritischen Zeichen geachtet wird. Der Projektbearbeiter übernimmt den dritten Korrekturdurchgang ebenfalls anhand der Originalquelle und strukturiert die Texte für eine spätere Veröffentlichung auf dem „Osteuropaserver“. Unter den gegenwärtigen finanziellen Bedingungen ist zu erwarten, daß bis zum Projektende (voraussichtlich im Februar 2003) eine aussagekräftige Auswahl von Dokumenten digitalisiert werden kann, die jedoch keineswegs einen umfassenden Anspruch erheben kann. Sollten die beantragten zusätzlichen Mittel durch die DFG genehmigt werden, kann diese Grundauswahl erheblich erweitert werden, zumal keine großen zusätzlichen Investitionen notwendig sind und die Einarbeitungs- und Orientierungsphase weitgehend abgeschlossen ist.

Das am UIM eingespielte Team hat seit Beginn der Projektlaufzeit zahlreiche Dokumente des 20. Jahrhunderts erfaßt und enge Verbindungen zu anderen Projekten innerhalb von ForOst geknüpft. Zudem werden mit anderen Instituten und Bibliotheken Gespräche über weitere Arten der Kooperation geführt. Denn der „Informationsserver“ versteht sich nicht als isoliertes Vorhaben, sondern ist dankbar für jede Art der Unterstützung und offen für eine Zusammenarbeit, die der Erfüllung des Forschungsvorhabens dient. Institutionen oder Personen, denen digitalisierte oder fotokopierte Quellen vorliegen, werden gebeten, diese dem Projekt in Kopie zur Verfügung zu stellen, damit sie durch den „Informationsserver“ des UIM bearbeitet, strukturiert und veröffentlicht werden können.²⁷ Sein Ziel, eine möglichst umfassende Quellensammlung zur Beziehungs- und Integrationsgeschichte im 20. Jahrhundert in Ostmittel- und Südosteuropa zu sein, kann dieser Server um so eher erreichen, je mehr Fachleute und Interessierte mit ihm kooperieren. Schließlich werden auch diese Gruppen von einem solchen umfassenden, zuverlässigen und zitierfähigen Medium profitieren und die Bedeutung der Geschichtswissenschaft für die Entscheidungsträger in Politik und Wirtschaft, aber auch für die interessierte Öffentlichkeit nachhaltig verdeutlichen können.

²⁷ Nähere Informationen und Kontakt über die Homepage des Ungarischen Instituts München: <http://www.ungarisches-institut.de>.