

Antrag

an die Deutsche Forschungsgemeinschaft

auf Gewährung einer Sachbeihilfe

Ausbau des Portals „Netzwerk SED-  
Archivgut“

zu einer Referenzanwendung für ein  
Archivportal Deutschland

Antragsteller:

Bundesarchiv

31. Juli 2006 (ergänzte Fassung vom 22. August 2006)

# **1 Allgemeine Angaben**

Antrag auf Gewährung einer Sachbeihilfe

## **1.1 Antragsteller**

Prof. Dr. Hartmut Weber  
Präsident des Bundesarchivs  
Potsdamer Str. 1  
56075 Koblenz  
0261 505200

## **1.2 Thema**

Ausbau des „Netzwerks SED-Archivgut“ zu einer Referenzanwendung für ein Archivportal Deutschland

## **1.3 Kennwort**

Ausbau Netzwerk SED-Archivgut

## **1.4 Förderungsbereich**

Überregionale Vorhaben der Archive

## **1.5 Voraussichtliche Gesamtdauer**

Das Projekt ist auf zwei Jahre konzipiert. Nach Abschluss des Projektes soll sich das Portal als Gemeinschaftsaufgabe der beteiligten Archive weitgehend selbst tragen. Das Bundesarchiv ist bereit, wenn die im Projektentwicklung vorgesehene Werkzeugentwicklung als Voraussetzung dezentraler Verantwortung für den Content ohne Zentralredaktion wie geplant stattfinden kann, für mindestens fünf Jahre den Portalbetrieb weiterhin zu gewährleisten und im Rahmen seines eigenen Internet-Auftritts die technische Infrastruktur zur Verfügung zu stellen und zu pflegen.

## **1.6 Antragszeitraum**

24 Monate

## **1.7 Termine**

1.10. 2006 – 30.9. 2008

## **1.8 Projektleitung**

Prof. Dr. Angelika Menne-Haritz,  
Vizepräsidentin, Bundesarchiv

## **1.9 Zusammenfassung**

Das Portal SED-Archivgut stellt seit April 2005 mit konstant hohen Zugriffszahlen Beschreibungen der Bestände früherer Parteiarchive der SED auf zentraler und regionaler Ebene bereit.

Das Portal wurde auf Initiative der Stiftung Archiv der Parteien und Massenorganisationen im Bundesarchiv in enger Zusammenarbeit mit den Landes- und Staatsarchiven der Länder Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen aufgebaut. In allen diesen Archiven werden Unterlagen der früheren Staatspartei der DDR archiviert. Inzwischen gibt es gemeinsame Bemühungen der beteiligten Archivverwaltungen, das Portal mit den Beschreibungen zu den zentral und dezentral archivierten FDGB-Beständen zu ergänzen.

Der Aufbau des Portals hat den am Netzwerk teilnehmenden Archiven die Möglichkeit gegeben, in einem Verbundfindmittel die vorhandenen Erschließungsleistungen weitgehend ohne neuen Erschließungsaufwand zusammen zu führen sowie gemeinsam und Institutionen übergreifend nutzen und durchsuchen zu lassen. Das Verbundfindmittel nutzt die Suchmaschine Midosasearch auf der Basis von Lucene. Dadurch kann es die vier Recherchemethoden: strukturierte Navigation, Volltextsuche, Blättern und Indexsuche kombiniert und zum jederzeitigen Wechsel anbieten. Kontaktinformationen und Benutzungsbedingungen sind außerdem an einer Stelle zugänglich. Neben den Beständeinformationen sind zur Erleichterung der Nutzung umfangreiche Hilfsmittel wie ein Glossar typischer Begriffe, eine Abkürzungsliste mit Auflösungen und Funktionärslisten gemeinsam erstellt worden. Das Portal wurde mit eigenen personellen und finanziellen Ressourcen ohne externe Fördermittel realisiert.

Die Webseite des Netzwerks SED-Archivgut entstand dadurch, dass die Archive ihre Erschließungsdaten in einer eigenen Software strukturiert oder mit der Software

MidosaXML aufbereitet haben. Die übrigen Angaben wurden in unterschiedlichen Formen, meist als Texte, an das Bundesarchiv geleitet. Soweit gewünscht, wurde den Archiven das Programm MidosaXML zur Verfügung gestellt, ansonsten wurden die Daten im Bundesarchiv mit diesem Werkzeug konvertiert. Die Integration der Daten in das Verbundfindmittel MidosaSEARCH wurde im Bundesarchiv in Berlin vorgenommen. Nach der Bereitstellung einer ersten Version der Daten im Verbundfindmittel wurden sie zunächst nur den beteiligten Archiven zur Einsicht bereitgestellt, so dass Veränderungen vorgenommen werden konnten. Alle beteiligten Archive lieferten ebenfalls Beiträge zum Glossar, zur Liste der Funktionäre wie zum Abkürzungsverzeichnis, die ebenfalls im Bundesarchiv in Berlin in die Präsentation integriert wurden. Dieses Verfahren ist für die weitere Pflege und den Ausbau des Portals nicht geeignet, weil es zu aufwändig ist. Es sollte, soweit es geht, automatisiert und mit geeigneten Werkzeugen rationeller gestaltet werden.

Zur dauerhaften Erhaltung und Aktualisierung der angebotenen Informationen ist die Entwicklung von Verfahren und Werkzeugen erforderlich, die den beteiligten Archiven die Pflege ihrer Daten erleichtern und die Abläufe soweit möglich automatisieren.

Dazu sind folgende Arbeiten von besonderer Bedeutung:

- Entwicklung von Konversions- und Editierwerkzeugen für die beteiligten Archive, mit denen sie vorhandene Daten aus ihren eigenen Systemen weitgehend automatisch in das gemeinsam genutzte Arbeitsformat bringen können.
- Entwicklung eines Harvesters, der die dezentral bereitgestellten Daten abholt. Alternativ dazu sollen für die Archive, die keinen Zugang zu einem Webserver haben, Verfahren der WebDAV-basierten Ablage der Daten auf dem zentralen Server beim Bundesarchiv angeboten werden. WebDAV (Web-based Distributed Authoring and Versioning) ist ein offener Standard zur Bereitstellung von Dateien im Internet. Dabei können Benutzer auf ihre Daten auf dem Server des Bundesarchivs wie auf eine Festplatte zugreifen. Für die Pflege der eigenen Erschließungsdaten und der allgemeinen Angaben zum Archiv kann ein Browser-Interface genutzt werden.
- Ausbau des zentralen Verbundfindmittels um eine dritte Ebene, um Archive, Beständeübersichten und Findbücher gemeinsam durchsuchen zu können. Bisher sind die Informationen zu den Archiven und zu den Beständen, also zwei Ebenen, in dem Portal durchsuchbar.

Mit solchen Verfahren und Werkzeugen kann das Portal dann auch als Ausgangspunkt für die Zusammenführung von weiteren Beständeinformationen genommen werden und weiteren Archiven geöffnet werden. Die neuen Entwicklungen sind die Grundlage für Strukturen, die für einen weiteren Ausbau des Netzwerks zu einem größeren Portal als erstes erforderlich wären. Es kann deshalb als Referenz genutzt werden, um Funktionen und Präsentationen zu testen und gegebenenfalls schrittweise für größere Datenmengen zu skalieren. Damit stellt es eine notwendige Referenzanwendung und logische Vorstufe zur sicheren Entwicklung eines deutschen Archivportals dar.

## 2 Ausgangslage, eigene Vorarbeiten

### 2.1 Ausgangslage

In ihrem Abschlussbericht „Die deutschen Archive in der Informationsgesellschaft – Standortbestimmung und Perspektiven“ hat sich die DFG-Arbeitsgruppe „Informationsmanagement der Archive“ im Herbst 2003 für die Entwicklung eines deutschen Archivportals vorzugsweise durch das Bundesarchiv ausgesprochen. Nach Abstimmung in der Konferenz der Archivverwaltungen des Bundes und der Länder ARK hat das Bundesarchiv im Frühjahr 2004 einen entsprechenden Antrag bei der DFG eingereicht. Diesem mit erheblichem Aufwand vorbereiteten Antrag und einer im Herbst 2004 eingereichten überarbeiteten Antragsfassung hat die DFG nicht entsprochen. Teile des Antrags, insbesondere methodische Fragen der Verwendung des Präsentationsstandards Encoded Archival Description EAD im Rahmen eines Verbundfindmittels für das Bundesarchiv konnten mit Hilfe der Andrew W. Mellon Foundation erfolgreich weiter bearbeitet werden.

Auf Anregung der Geschäftsstelle der DFG fand am 10. Januar 2006 ein Gespräch zwischen einer Delegation des Unterausschusses Informationsmanagement und der Leiterin des Bereichs Wissenschaftliche Informationssysteme 4 beim Präsidenten des Bundesarchivs statt mit dem Ziel, Möglichkeiten der Fortführung des Projekts Archivportal Deutschland auszuloten. Beide Seiten haben sich dabei übereinstimmend für eine stufenweise Realisierung eines deutschen Archivportals ausgesprochen und der Entwicklung des Archiv und Bestände übergreifenden Verbundfindmittels als Kernstück des Portals Priorität eingeräumt. Das Bundesarchiv hat dabei vorgeschlagen, in einem ersten Schritt das mit Eigenmitteln realisierte Verbundfindmittel des Netzwerks SED-Archivgut als Referenzmodell für ein skalierbares Verbundfindmittel des deutschen Archivportals auszubauen. Von Seiten der DFG wurde empfohlen, einen entsprechenden Antrag vorzubereiten und einzureichen.

Das Bundesarchiv hat am 12. Januar 2006 die Mitglieder der Archivreferentenkonferenz Ost (Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt, Thüringen und Bund) sowie nach schriftlicher Vorabinformation am 21./22. März die Archivreferentenkonferenz über die neue Konzeption informiert und diese mit den Ländern erörtert. Auf Anregung der Archivreferentenkonferenz lud das Bundesarchiv für den 11. Mai 2006 zur einer Informationsveranstaltung ein, auf der die vorhandenen Vorarbeiten, die in den Portalaufbau integriert werden können, vorgestellt wurden. Es nahmen Vertreter von 14 Archiven teil. Ein Bericht über die Veranstaltung und ihre Ergebnisse sowie die Präsentationen, die bei der Veranstaltung vorgestellt wurden, sind unter: <http://www.bundesarchiv.de/aktuelles/fachinformation/00049/index.html> auf der Website des Bundesarchiv einsehbar. Die Teilnehmer der Veranstaltung waren sich einig, dass der Aufbau eines deutschen Archivportals dringlich ist, und dass der Weg, den das Bundesarchiv mit seinen Entwicklungen eingeschlagen hat, also der Einsatz internationaler, offener Standards und die Entwicklung von Anwendungen nachnutzbarer Software, für den Portalaufbau nutzbar gemacht werden sollte. Als wichtiger Punkt bei den Schritten zum Portalaufbau wurde die Entwicklung von frei nachnutzbaren Werkzeugen für die Archive angesehen, die es ihnen erlauben, ihre

eigenen Erschließungsdaten ohne Zusatzaufwand in ein zentrales Verbundfindmittel zu integrieren.

In Europa gibt es inzwischen leistungsfähige zentrale Archivportale in mehreren Ländern neben mehreren regionalen oder themenbezogenen Portalen. Vorreiter sind Italien und Spanien mit einer vergleichbaren Architektur ihrer Portale. Das spanisch-iberoamerikanische Censo-Guía bietet den integrierten Zugriff auf Informationen zu Beständen aus 40.000 spanischsprachigen Archiven weltweit. Das italienische Portal SIAS erlaubt eine Recherche über Bestände aus 134 Staatsarchiven und das Portal SIUSA bietet einen integrierten Zugang zu nicht staatlichen Archiven.

Alle drei Portale bieten als Einstieg drei miteinander verknüpfte Bereiche für die übergreifende Suche, nämlich den Bereich der Beschreibungen von ganzen Archivgutbeständen in Findbüchern, einen Bereich mit Informationen über die Herkunftsstellen, ähnlich den Beständeübersichten und schließlich einen Bereich mit Angaben zu den Archiveinrichtungen an. In beiden Ländern werden die Daten dazu in den internationalen Standards EAD (Encoded Archival Description), EAC (Encoded Archival Context) und EAG (Encoded Archival Guide) strukturiert und verwaltet. EAD ist der aus Amerika stammende Standard für die Struktur von Findbüchern und wird für die Verzeichnungsangaben und Findbücher verwendet. EAC wurde von einer internationalen Arbeitsgruppe im Rahmen des LEAF-Projektes gemeinsam von Archivaren und Bibliothekaren entwickelt als eine archivisch erweiterte Struktur für normierte Ansetzungsformen, die mit Erläuterungen versehen werden können. Es wird in den Portalen für den Sucheinstieg über die Herkunftsstellen eingesetzt. EAG wurde in Spanien speziell für den Censo-Guía entwickelt und wird zunehmend auch in anderen Ländern für Informationen zu den Archivinstitutionen, wie etwa Adressen und Öffnungszeiten, eingesetzt. In den Portalen wird es für die Informationen zu den Archiveinrichtungen, ihre Adressen, Öffnungszeiten etc. verwendet. EAD, EAC und EAG sind XML-DTDs oder -Schemata. Die Daten werden oder wurden in der Praxis aus Datenbank Anwendungen, die ISAD(G) oder ISAAR(CPF) nutzten, in die XML-Formate konvertiert und in dieser Form für die Suche und die Präsentation in HTML genutzt. Das spanischsprachige Portal bietet die Möglichkeit, mit einem Klick auf ein kleines Symbol sich direkt auch die XML-Dateien in allen drei Standards anzeigen zu lassen.

Ähnlich aufgebaut sind Portale in Frankreich und England. Frankreich hat schon vor drei Jahren mit Kanada zusammen das Portal Canada-France mit Materialien zur gemeinsamen Geschichte, angereichert mit Images von Dokumenten und vielen Abbildungen, aufgebaut, das aktuell erneuert wird. Dazu wird EAD genutzt. Das Nationalarchiv in Großbritannien hat mit 4 Mill £ Förder- und Lottogeldern das Portal A2A (Access to Archives) mit knapp 9 Mill. Verzeichnungseinheiten aus über 400 Archiven aufgebaut. Ebenso wie bei ArchivesHub, dem zentralen Zugriff zu den Universitätsarchiven in Großbritannien, wurde zunächst ISDA(G) für die Beständebeschreibungen genutzt. Inzwischen ist die Anwendung auf EAD umgestellt. EAC wird weniger genutzt und EAG wird neuerdings für ARCHON, die online-Version des National Registers of Manuscripts, eingesetzt.

Außerdem haben erste Schritte für den Aufbau eines europäischen Archivportals begonnen. Das European Board of National Archivists (EBNA) hat in seinem Bericht für die europäische Kommission den Aufbau eines europäischen Archivportals

empfohlen, dessen Architektur den EAD-Standard nutzt. Nach der Verabschiedung des Berichts wird im Herbst 2006 die Arbeitsgruppe für den Aufbau des Portals konstituiert. Die Federführung liegt bei Spanien, das inzwischen in Europa bei der Portalentwicklung im Archivbereich den größten Vorsprung hat.

In Deutschland wurden in den letzten Jahren mehrere Regionalportale aufgebaut. Ihre Architektur beruht meist auf der Nutzung einer zentralen Datenbank, in die dezentrale Erschließungsinformationen importiert werden und in denen eine webbasierte Suche durchgeführt wird. Einige größere Portale (archive-nrw.de; hadis) nutzen inzwischen XML-Technologie für den Aufbau der Präsentation der Rechercheergebnisse in ihren Kontexten oder bereiten diesen Schritt vor. Bei dem neuen Angebot aus Rheinland-Pfalz mit einer Datenbanksuche wie auch bei der Präsentation der Online-Findbücher in Baden-Württemberg mit einem File-System wird die strukturierte Präsentation der Findmittel in Kombination mit der Volltextsuche eingesetzt, die eine Navigation unter Nutzung der Zusammenhänge von Erschließungsangaben erlaubt. Das BAM-Portal baut als erste deutsche Initiative dieser Art einen gemeinsamen, Institutionen übergreifenden Zugang zu Archiv-, Bibliotheks- und Museumsbeständen auf. Nach anderen Ansätzen setzt es inzwischen auch auf EAD als Format für neu zu integrierende Daten.

Andere Portale wie clio-online nutzen Metasuchmaschinen, mit denen bei einer Anfrage Suchprozesse in verteilten Datenbanken angestoßen werden und die Ergebnisse in einem einheitlichen, zentralen Format angezeigt werden. Der Einsatz von Metasuchmaschinen hat offenbar nicht die oft damit verbundenen Erwartungen realisieren können. Nach den Erfahrungen des Projektes Clio-Online ist der Zugriff auf heterogene Datenbestände hochkomplex. Änderungen in der Struktur der abgefragten Datenbanken sind schwierig nachvollziehbar und die formal vereinheitlichende Darstellung verwischt ursprünglich intendierte Bedeutungen. Aus den gleichen Gründen hat sich die Arbeitsgruppe aus Staatsbibliothek und Bundesarchiv zur Form und Zielstellung eines gemeinsamen Nachweises von Nachlässen aus den Datenbanken Kalliope und ZDN (Zentrale Nachlassdatenbank, gehostet beim Bundesarchiv) gegen den Einsatz einer Metasuche ausgesprochen. (Vgl. Positionspapier für die DFG vom 9.6.2006). Es wird stattdessen davon ausgegangen, dass mit den XML-Standards EAC und EAD die passenden Formate für einen Datenaustausch vorliegen.

Das Internetangebot des Netzwerks SED-Archivgut besteht zur Zeit aus den Erschließungsangaben auf Bestandsebene sowie Informationen zu den 13 beteiligten Archiven und den von ihnen zur Nutzung angebotenen Materialien zusammen mit Glossar, Abkürzungsverzeichnis und Funktionärslisten. Im Zentrum steht das auf der Basis der Suchmaschine MidosaSEARCH arbeitende Verbundfindmittel, das eine übergreifende Recherche in den Erschließungsangaben aller dreizehn Archive erlaubt und die Ergebnisse innerhalb der strukturierten Präsentation anzeigt. Die Trefferlisten werden expandierbar strukturiert und erlauben den direkten Zugriff auf die Fundstellen innerhalb des vollständigen Textes der Findbücher. Es kann von Fundstelle zu Fundstelle gesprungen werden, ohne zwischendurch zur Trefferliste zurückkehren zu müssen, während zwischendurch die Blätterfunktion oder die Gliederung zur Navigation in anderen Bereichen des Findbuchs genutzt werden kann.

Das Verbundfindmittel im Portal SED-Archivgut benutzt zwei Ebenen. Auf der oberen Ebene sind die Archive nach Bundesländern aufgelistet. Auf der Ebene darunter befinden sich damit verknüpft die Beständebeschreibungen als Auszüge aus den Beständeübersichten der Archive mit ihrer jeweiligen tektonischen Gliederung. Auf jeder Ebene können Bereiche für die übergreifende Volltextsuche, also z.B. einzelne Archive oder bestimmte Bestandsgruppen, ausgewählt und zusammengestellt werden. Wenn keine Auswahl definiert wurde, wird über alles gesucht.

Das Verbundmittel auf der Basis von MidosaSEARCH stellt eine innovative Kombination aus Volltextsuche, strukturierter Navigation und Blättern mit ähnlichen Eigenschaften wie e-books dar. Wie bei einem e-book können die Gliederungsgruppen als Kapitel vom Navigationsframe als Inhaltsverzeichnis aus erreicht werden und ein Index hilft bei der Suche von Begriffen, die noch nicht bekannt sind.

Die Trefferanzeige der Volltextsuche in MidosaSEARCH erlaubt das gezielte Ansteuern der Fundstellen in den Findbüchern und das Weiterblättern zur nächsten Fundstelle im Kontext. Jederzeitiger Wechsel zwischen den Rechercheverfahren ist möglich. Wenn man etwa mit Hilfe der Volltextsuche in die Recherche einsteigt, kann man zunächst von der Fundstelle aus in der Struktur weiter navigieren oder im Findbuch vor und zurück blättern, und dann zur nächsten Fundstelle springen.

Ein wichtiger Nebeneffekt entsteht durch die strukturierte Darstellung der Gliederung. Mit ihrer Hilfe zeigt das Verbundfindmittel die Organisationsstruktur der SED und gibt so Besuchern des Portals direkt die Informationen darüber, welche Grundorganisationen der SED wo existiert haben und wie sie bezeichnet wurden.

Für das SED-Archivgut-Netzwerk ist eine eigene Installation der Suchmaschine MidosaSEARCH eingerichtet worden. Sie wird regelmäßig im Zuge der Erweiterung und Weiterentwicklung von MidosaSEARCH, die im Bundesarchiv vorgenommen wird, aktualisiert.

Das Verbundfindmittel verbindet in einer innovativen Kombination vier verschiedene Rechercheverfahren, nämlich die Volltextsuche, die strukturierte Navigation, das Durchblättern des kompletten Textes und die Nutzung eines Index mit Links zur entsprechenden Verzeichnungseinheit. Diese Kombination bewährt sich seit längerem in den Online-Findbüchern des Bundarchivs und seit einem Jahr in der eigenen Installation der Suchmaschine MidosaSEARCH, die, wie auch das Portal SED-Archivgut, ständig anwachsende hohe Nutzungsfrequenzen zeitigt.

Dem Portal SED-Archivgut liegt eine ähnliche Architektur zugrunde wie sie bei den Entwicklungen in anderen Ländern Europas in den letzten zwei Jahren festzustellen ist. Außerdem wird von der Gruppe der Europäischen Nationalarchivare in ihrer Empfehlung zum Aufbau eines europäischen Portals das gleiche Verbundfindmittel mit den kombinierten Recherchemethoden und der Verbindung mehrerer Ebenen von Erschließungsinformationen vorgeschlagen. Das Portal SED-Archivgut kann deshalb mit seiner Architektur sehr gut eine Vorbereitung für den deutschen Beitrag zum europäischen Portal darstellen und als Schnittstelle für die Integration von Informationen über deutsche Archive in internationale Verbünde fungieren.

Das Portal benutzt eine Methodik, die den Ansätzen der anderen europäischen Länder entspricht. Die Entscheidung für diese Methodik wurde durch die dabei mögliche Nutzung von XML erleichtert, weil damit die Arbeiten an dem Portal relativ wenig Aufwand erforderten und die Konvertierung der vorhandenen Daten in ein einheitliches Format für das Verbundfindmittel mit der vorhandenen Software durchgeführt werden konnte. Für die gemeinsam genutzten Erschließungsangaben wird ein XML-Format eingesetzt (MidosaXML), das weitgehend der deutschen Übersetzung von EAD entspricht und mithilfe der Software MidosaXML auch in EAD umgewandelt werden kann. EAC oder EAG wurden bisher nicht genutzt. Es ist sinnvoll, EAC für die Beschreibungen der Herkunftsstellen und EAG für die Informationen zu den Archiven einzusetzen, weil sie den Vorteil eines bereits vorhandenen einheitlichen Formats haben und zudem international austauschbar sind. Mit dem XML-Format konnten die Daten in das Verbundfindmittel integriert werden und können nun mehrdimensional bestands- und archivübergreifend durchsucht werden.

Neben MidosaSEARCH und neuen Entwicklungen des BAM-Portals, bei dem das Bundesarchiv Partner ist, und das inzwischen EAD als Importformat verwendet, gibt es in Deutschland ein wachsendes Interesse an EAD. Die Erweiterung des Portals für eine größere Zahl von Archiven und Bestände sollte schon wegen der internationalen Kompatibilität und des möglichen Austauschs mit einem europäischen Archivportal oder anderen Anbietern von gemeinsamen Zugriffspunkten für die zur übergreifenden Recherche erforderlichen Daten das Format EAD einsetzen. Das Bundesarchiv bemüht sich aus diesem Grund um eine Erleichterung des Zugangs zu Informationen über EAD für deutschsprachige Archivarinnen und Archivare. Im Rahmen des Projektes <daofind> wurden neben Grundinformationen und Listen wichtiger Links Übersetzungen der wichtigsten Dokumente einschließlich der Application Guidelines und der Tag-Library erstellt und auf der Projekt-Seite [www.daofind.de](http://www.daofind.de) zum Download angeboten.

Das Bundesarchiv richtet im April 2007 im Rahmen der deutschen Präsidentschaft der EU nach London und Paris die 3. europäische Konferenz zu EAD, EAC und METS unter dem Titel „Internationale Standards für digitales Archivgut“ in Berlin aus. (Vgl. [www.instada.eu](http://www.instada.eu)) Daran werden auch Referenten aus den Editorial Boards der drei Standards EAD, EAC und METS mitwirken. Hier könnten die deutschen Schritte zur Entwicklung eines Portals vorgestellt und in die internationale Diskussion eingebunden werden.

Außerdem ist das Bundesarchiv Partner der National Agency for Germany im EU-Projekt Michael+ und kann darauf achten, dass die deutschen archivischen Erfahrungen in den Aufbau des europäischen Nachweissystems von Kulturgut eingebracht werden.

## **2.2 Eigene Vorarbeiten**

Das Bundesarchiv bringt das Verbundfindmittel, das Kernstück des Portals SED-Archivgut ist, als eigene Vorarbeit ein. Besondere Merkmale des Verbundfindmittels auf der Basis der Suchmaschine MidosaSEARCH, wie sie für das Portal zum SED-Archivgut eingesetzt wird, sind:

- Übergreifende Recherche in allen Erschließungsangaben des Portals;

- Anzeige der Treffer einer Volltextsuche als Fundstelle im vollständigen Text mit der Möglichkeit direkter Navigation zur nächsten Fundstelle;
- Beliebiger Wechsel zwischen Volltextsuche, Navigation in der Struktur, dem Durchblättern der Beständebeschreibungen und einer Index-gestützten Suche;
- Möglichkeit der Auswahl der gemeinsam zu durchsuchenden Archive und Bestände.

Außerdem wird die fertig aufgebaute bisherige Struktur der Präsentation mit den Bereichen für Informationen über die Archive und den unterstützenden Informationen wie Glossar, Abkürzungsverzeichnis und Funktionsliste eingebracht. Layout und Struktur können einem vergrößerten Informationsangebot angepasst werden.

Bisherige Funktionen und Vorarbeiten für das Portal SED-Archivgut sowie die dabei und in anderen Projekten vorgenommenen Entwicklungsarbeiten werden in das Projekt eingebracht. Außerdem ist das Bundesarchiv bereit, weitere Entwicklungen für seine eigenen Internet-Angebote für die Erweiterung des Portals bereitzustellen, die sich in der Entwicklung befinden. Dazu gehören:

- Anpassung der Performance von MidosaSEARCH an größere Datenmengen.

Das Verbundfindmittel auf der Basis der Suchmaschine MidosaSEARCH wird im Bundesarchiv als Präsentationswerkzeug für alle Bestände eingesetzt. Dafür wird der Auftrag zur Erweiterung von MidosaSEARCH erteilt, um die Suchmaschine für alle Bestände einschließlich der aus der Oracle-basierten Datenbank BASYS exportierten Erschließungsinformationen übergreifend einzusetzen. Das bedeutet die Ausweitung der Kapazität auf über 6000 Bestände und 1,5 Mill Verzeichnungseinheiten. Dabei wird jedoch noch nicht die dritte Ebene realisiert, weil für das Bundesarchiv allein zwei Ebenen, nämlich für die Beständeübersicht und die damit verknüpften Findbücher ausreichend sind und diese Erweiterung deshalb nicht aus Eigenmitteln finanziert werden kann.

- Werkzeuge aus dem Projekt <daofind>

Im Rahmen des Projektes <daofind> (Digitalisiertes Archivgut in Online-Findmitteln) wurden im Bundesarchiv mit Eigenmitteln und mit Förderung durch die Andrew W. Mellon-Foundation prototypische Editoren und Pilotanwendungen für die Präsentation von Image-Digitalisaten aus Archivgut entwickelt. Die Editoren für Texte und Metadaten nutzen die XML-Schemata der drei Standards EAD, EAC und METS (METS=Metadata Encoding and Transmission Standard). Es handelt sich dabei um fachsprachliche Dokumenteneditoren für EAD-, EAC- und METS-Dateien. Die Editoren bieten Abschnitte und Elemente der entsprechenden Dokumente, also etwa von Findbüchern, für die Eingabe von Inhalten an und validieren das so angelegte oder überarbeitete Dokument gegen das entsprechende Schema. Sie bieten eine komfortable Möglichkeit, XML-Exporte aus Datenbanken zu ergänzen oder anzupassen sowie neue Dateneingaben mit der gewohnten Fachterminologie zu unterstützen. Der METS-Editor aus diesem Werkzeugset kann für die Einbindung von Digitalisaten in Online-Findbüchern genutzt werden, wobei wiederholte Arbeitsvorgänge automatisiert werden können.

Die Editoren sind auf der Basis der OpenSource Entwicklungsumgebung Eclipse in Java programmiert und können deshalb plattformunabhängig eingesetzt werden. Sie können als Rohform für maßgeschneiderte Konversions- oder Erfassungswerkzeuge für die beteiligten Archive genutzt werden.

- Bereitstellung eines WebDAV-Zugangs zum zentralen Server alternativ zum Harvesting dezentral bereitgestellter Daten.

Das Bundesarchiv kann für die beteiligten Archive einen externen WebDAV-Zugang für den direkten Zugriff auf ihre eigenen Daten, die auf dem zentralen Server liegen, bereithalten. Ähnliche Anwendungen werden im Bundesarchiv bereits erfolgreich für die Pflege der Zentralen Nachlassdatenbank (ZDN) mit 26.000 Nachweisen weltweit und für das Katastrophenschutzregister NORA sowie für den Findmittel-Guide im Rahmen von Clio-online eingesetzt. Die WebDAV-Möglichkeit soll denjenigen Archiven angeboten werden, die keinen eigenen Zugang zu einem Webserver haben, auf dem sie ihre Daten für einen Harvester bereithalten können.

Das Bundesarchiv ist außerdem bereit, für mindestens 5 Jahre das Portal auf seinen Servern zu hosten und die Hardware und Software sowie die erforderliche Systemadministration für das Portal bereitzustellen. Soweit dazu allerdings Erweiterungen von Software-Lizenzen erforderlich sind, müssten diese aus den Fördermitteln bereitgestellt werden. Während der garantierten Betriebszeit von 5 Jahren wird sich das Bundesarchiv um ein Betriebsmodell bemühen, das die Verstetigung des Portals gewährleistet.

### **3 Ziele und Arbeitsprogramm**

#### **3.1 Ziele**

Ziel des Projektes ist es, das bestehende Portal und seine Infrastruktur um die zusätzlichen Funktionen zu ergänzen, die seine regelmäßige Aktualisierung und dezentrale Pflege ermöglichen. Dabei geht es vorrangig um die Nachhaltigkeit der bisherigen und zukünftigen Arbeiten am Portal, wobei der Ausbau der Darstellung selbst und der angebotenen Funktionen einen geringeren Aufwand als die Entwicklungen für die erforderliche Infrastruktur erfordern. Deshalb liegt der Schwerpunkt des Projektes beim Ausbau des Informationsangebots mit der Integration der Findbücher in das Verbundfindmittel und der Bereitstellung von Instrumenten zur Pflege der Inhalte.

Der Stand der bisherigen Entwicklungen für das Portal SED-Archivgut ist für den aktuellen Umfang und für ein weitgehend statisches Angebot ausreichend. Vor allem unter dem Aspekt einer Ausweitung für weitere Archive und weitere Bestände ist jedoch Erweiterungsbedarf und die Notwendigkeit der Unterstützung für die Bereitstellung des Inhaltes ebenso wie die Integration einer dritten Ebene in das Verbundfindmittel für die Findbücher vorhanden.

Mit dem Projekt ist die Ausweitung der Kapazität der Portals SED-Archivgut bei grundsätzlicher Beibehaltung seiner Funktionalitäten geplant. Damit kann schrittweise eine Struktur für das Verbundfindmittel als Kernstück eines deutschen Archivportals vorbereitet werden. Es soll nach dem Ausbau im Rahmen des Projektes für beliebige Arten von Archiven und Archivgut offen stehen. Deshalb ist als letztes Arbeitspaket des Projektes die Prüfung auf die Verwendbarkeit oder weiteren Erweiterungsbedarf für ein Archivportal-D vorgesehen.

Wichtiges Ziel bei der Entwicklung von neuen Werkzeugen und Instrumenten ist das Prinzip, sich wiederholende Arbeitsschritte zu automatisieren und gleichzeitig den Spielraum für erforderliche oder gewünschte Fachentscheidungen bereitzustellen. Den Archiven, die Inhalte für das Portal bereitstellen wollen, soll es ermöglicht werden, ohne zusätzliche Aufwände die erforderlichen Arbeiten in die Tagesarbeit ihrer Mitarbeiter zu integrieren.

### **3.2 Arbeitsprogramm**

Für die Erweiterung des Portals sind Arbeiten in folgenden Bereichen erforderlich:

1. Erweiterung der Kapazität und Leistungsfähigkeit des Verbundfindmittels durch Weiterentwicklung der Suchmaschine über das für den Einsatz im Bundesarchiv erforderliche Maß hinaus mit Orientierung auf die Integration unterschiedlicher Archive, verschiedenartiger Bestände und unterschiedlicher Ausgangsformate von Erschließungsdaten. Dazu gehört auch eine noch weitere Skalierung der Suchmaschine für größere Datenmengen über 1,5 Mill Verzeichnungseinheiten hinaus sowie die Anpassung der Performance für große Datenmengen.
2. Erweiterung der Suchmaschine um eine dritte Ebene für die Integration von Findbüchern neben den Informationen zu den Archiven und den Beschreibungen der Bestände, also den Beständeübersichten. Auf den beiden bisher angebotenen Ebenen können Archive und darunter Bestände für eine übergreifende Suche ausgewählt werden. Sie sollten um eine weitere Ebene für die Recherche in Findbüchern erweitert werden. Das betrifft auch die Aggregation in der Anzeige der Treffer. Sie werden in einer aufklappbaren Struktur angeboten. Damit wird eine sinnvolle Reduktion der Anzahl angezeigter Treffer erreicht, ohne dass die Trefferzahl künstlich beschränkt wird.
3. Anpassung der vorhandenen Editoren aus dem Daofind-Projekt des Bundesarchivs als Werkzeuge für die dezentrale Vorbereitung der Daten und deren eventuell erforderliche Konversionen. Den Archiven soll die Arbeit bei der Aufbereitung eigener Daten für die Integration in das Portal durch angepasste Werkzeuge erleichtert werden. Deshalb gehört zu den Erweiterungen des Portals der Aufbau eines Werkzeugkastens für die beteiligten Archive, mit dem sie ihre Erschließungsdaten aus eigenen Datenformaten, soweit gewünscht und erforderlich, in das gemeinsame Austauschformat bringen können, ihre Grundinformationen zu ihrer Einrichtung in die entsprechende Datenbank einpflegen oder auch e-Learning-Tools für potentielle Benutzer erstellen können. Dazu gehört auch die Umstellung der

Suchmaschine MiodsaSEARCH von den deutschsprachigen Tags von MidosaXML auf die englischsprachigen Tags von EAD.

4. Entwicklung von Verfahren und Instrumenten für die Bereitstellung der dezentralen Daten und ihre Integration in das Verbundfindmittel. Dazu wird ein Harvester benötigt, der evtl. aus anderen Projekten übernommen und angepasst werden kann, und mit dem die dezentral bereitgestellten Daten vor Ort indexiert und im erforderlichen Umfang kopiert werden.
5. Für die Archive, die ihre Daten nicht dezentral bereitstellen wollen oder können, soll ein WebDAV-Zugang auf dem Server des Bundesarchiv eingerichtet werden, wo die Archive ihre eigenen Daten so bearbeiten, speichern und freigeben können wie auf den eigenen Speichern. Dazu soll ein browserbasierter Zugang angeboten werden.
6. Ausbau der Verlinkungsfähigkeit auf jeder Ebene für die Verknüpfungen zu den eigenen Angeboten der Archive. Bisher wird von der Einstiegsseite des Portals aus ein direkter Link zu den Archiven angeboten. Nach Möglichkeit sollte aus der übergreifenden Suche in den Findmittelinformationen an beliebiger, von den Archiven wählbarer Ebene, auf deren eigene Präsentation verlinkt werden.
7. Schließlich sollte das Verbundfindmittel angesichts der begonnenen Diskussion um die Bereitstellung von Digitalisaten aus den Beständen der Archive um eine Funktion für die Präsentation von Images innerhalb der Findbücher erweitert werden. Dazu liegen ebenfalls Vorarbeiten aus dem Daofind-Projekt des Bundesarchivs vor. Sie umfassen ein Werkzeug für eine Integration von Images aus den Beständen in Online-Findmitteln und eine speziell für Schriftgut entwickelte Präsentationsform, bei der die Abbildungen einer Verzeichnungseinheit zu einem digitalen archivischen Objekt zusammengefasst werden und die Abbildungen mit Einfluss auf die Struktur der Akte auf einer Orientierungsoberfläche jeweils mit dem oberen Drittel der Seite gezeigt werden.
8. Zum Abschluss sollte die Skalierbarkeit des Portals geprüft werden, um festzustellen welche zusätzlichen Entwicklung für den weiteren Ausbau als deutsches Archivportal erforderlich sind, damit ein weiterer Förderantrag fundiert vorbereitet werden kann. Dabei reicht es sicherlich aus, in einem ersten Ausbauschnitt von einer Skalierung für eine Menge von etwa 4 Mill. Verzeichnungseinheiten aus 400 Archiven auszugehen.

### **3.3 Arbeitsplan**

Die für den Ausbau erforderlichen Entwicklungen sollten mit den folgenden Arbeitspaketen umgesetzt werden:

#### **Teilprojekt 1: Feinplanung und Auftaktarbeiten**

**Arbeitsbereich 1:** Feinplanung / Mengengerüst / Ausschreibungen

Der Arbeitsbereich 1 unterteilt sich in die folgenden Arbeitspakete:

AP 1.1: Konkretisierung des Projektkonzepts

AP 1.2: Vorbereitung der Beschaffung für die Softwareentwicklung

AP 1.3: Einrichten einer Projektwebsite

Ergebnis aus Arbeitsbereich 1: Schriftlich fixiertes Konzept zur den fachlichen und technischen Anforderungen einschließlich eines Sicherheitskonzepts und eines Mengengerüsts. Die Arbeiten finden zeitversetzt statt. Erforderliche Kapazität: 2 Personenmonate

**Arbeitsbereich 2:** Vorbereitung und Bereitstellung der technischen Infrastruktur  
Der Arbeitsbereich 2 unterteilt sich in die folgenden Arbeitspakete:

AP 2.1: Festlegungen der konkreten Anforderungen an die technische Ausstattung und die Nutzung der internen Netze und Strukturen sowohl für die Kommunikation nach innen und außen wie für die Übernahme und Integration der externen Daten, gemeinsam mit dem IT-Referat des Bundesarchivs.

AP 2.2: Festlegung der Anforderungen an Metadaten und Dokumenten durch die IT-Systemverwaltung für die Integration der Daten und ihre dauerhafte Verwaltung und Einrichtung des Servers.

Ergebnis aus Arbeitsbereich 2: Schriftlicher Katalog der Anforderungen der IT-Seite. Die Arbeiten finden zeitversetzt statt. Erforderliche Kapazität: 2 Personenmonate

**Arbeitsbereich 3:** Abstimmungen mit den bisher am Portal beteiligten Archiven über Erweiterungswünsche und Anforderungen aus der eigenen Erfahrung im bisherigen Umgang mit dem Portal.

AP 3.1: Aufstellung eines Katalogs relevanter Fragen und Vorstellungen und ihre Priorisierung

AP 3.2 Veranstaltung eines Workshops auf der Grundlage schriftlich übermittelter Antworten nach einer Auswertung bereits vorher stattgefundenen Treffen und eines auf dieser Grundlage erarbeiteten Fragebogens.

Ergebnis aus Arbeitsbereich 3: Schriftliche Planung der Erweiterungen und Veränderungen einschließlich der Prioritätensetzung

Die Arbeiten finden zeitversetzt statt. Erforderliche Kapazität: 6 Personenmonate

### **Ergebnisse des Teilprojektes 1:**

In diesem Teilprojekt werden die erforderlichen Vorbereitungsarbeiten für den Einstieg in die konkrete Softwareentwicklung erledigt. Die funktionellen Anforderungen sind definiert und die technische Infrastruktur des Bundesarchivs ist auf die im Projekt entstehenden Erfordernisse vorbereitet. Die bisher schon am Aufbau des Portals SED-Archivgut beteiligten Archive haben die Gelegenheit, ihre bisherigen Erfahrungen einzubringen und für die Entwicklungen zum Tragen kommen zu lassen.

**Gesamtkapazität:** 10 Personenmonate

## Teilprojekt 2: Realisierung der angestrebten Software-Entwicklungen

**Arbeitsbereich 4:** Ausbau der Kapazität und Leistungsfähigkeit des Verbundfindmittels durch Weiterentwicklung der Suchmaschine

AP 4.1: Erweiterung der Kapazität und Anpassung der Performance. Die Arbeiten werden extern erledigt. Es muss eine interne, auch telefonische Betreuung sichergestellt sein und es sind regelmäßige Tests erforderlich. Dazu muss ein Testscenario aufgebaut und eine Checkliste erstellt werden. Die internen Arbeiten werden sich zu bestimmten Zeiten häufen, sind aber ansonsten parallel zu anderen Arbeiten erledigen.

AP 4.2: Erweiterung um die dritte Ebene für die Integration der Findbücher. Auch diese Arbeiten werden extern erledigt, so dass intern ebenfalls die Betreuung sichergestellt sein muss. Regelmäßige Tests sind erforderlich. Dazu muss ein Testscenario aufgebaut und eine Checkliste erstellt werden. Die internen Arbeiten werden sich zu bestimmten Zeiten häufen, sind aber ansonsten parallel zu anderen Arbeiten zu erledigen.

Ergebnis aus dem Arbeitsbereich 4: Die Suchmaschine ist so weit ausgebaut, dass sie für die Einspielung weiterer Bestände und größerer Mengen an Verzeichnungseinheiten in das Verbundfindmittel vorbereitet ist. Erforderliche Kapazität: 10 Personenmonate

**Arbeitsbereich 5:** Erstellung und Anpassung der Werkzeuge und Verfahren für die beteiligten Archive zur Bereitstellung und Pflege ihrer Daten

AP 5.1: Anpassung der vorhandenen Editoren aus dem daofind-Projekt für die dezentrale Vorbereitung der Daten mit den notwendigen Automatisierungsfunktionen einschließlich der Mappings für die Konversion in das einheitliche Arbeitsformat. Die Mappings werden mit den beteiligten Archiven an Hand ihrer intern genutzten Datenstrukturen erstellt. Dabei werden sich viele Arbeiten wiederholen. Außerdem kann auf vorhandene Mappings wie MidosaXML <-> EAD und Austauschformat (SAFT) <-> EAD zurückgegriffen werden. Die Anpassung der Editoren erfolgt extern, muss aber intern betreut werden. Dazu sind Tests mit Daten aus den jeweiligen Archiven erforderlich. Zu diesen Entwicklungen gehört die Anpassung des EAD-Editors für Findbücher und Beständeübersichten und des EAC-Editors für die Angaben zu den Herkunftsstellen, die von der Beständeübersicht aus aufgerufen werden können. Außerdem gehört dazu die völlig neue Programmierung des EAG-Editors für die Informationen zu den Archiven in einem mit den Plänen für das europäische Portal kompatiblen Format.

AP 5.2: Entwicklung der Verfahren für die Übermittlung oder Bereitstellung der Daten einschließlich der Prüfung, ob ein vorhandener Harvester übernommen und angepasst werden kann. Dazu muss zunächst festgestellt werden, welche vergleichbaren Harvester bisher entwickelt wurden und zur Verfügung gestellt werden könnten. Dann ist der erforderliche Anpassungsaufwand abzuschätzen und mit dem Aufwand einer Neuentwicklung sowie den Kosten für einen dauerhaften Einsatz und später notwendige Updates abzuwägen. Danach kann die eigene Entwicklungsarbeit beginnen, die ebenfalls extern vergeben werden sollte.

AP 5.3: Erstellung und Anpassung des WebDAV-Zugangs zu einem Server des Bundesarchivs einschließlich einer browsergestützten Nutzungsoberfläche für Ablage und Pflege der in das Portal einbeziehenden oder bereits einbezogenen Daten. Erfahrungen mit WebDAV-Arbeit liegen im Bundesarchiv bereits vor, da MidosaXML WebDAV-fähig ist, und die gemeinsame Arbeit an Findbüchern im internen Netz diese Funktion nutzt. Außerdem ist bereits seit einiger Zeit erfolgreich eine browsergestützte Eingabe und Pflege von Daten sowohl bei der zentralen Nachlassdatenbank (ZND) sowie bei dem Notfallregister für Archive NORA eingesetzt.

Ergebnis aus Arbeitsbereich 5: Die Verfahren und Werkzeuge für die Integration neuerer Daten in größeren Mengen stehen bereit, so dass mit ihrer probeweisen Einbeziehung begonnen werden kann. Die Arbeiten können z.T. parallel durchgeführt werden. Erforderliche Kapazität: 12 Personenmonate

**Arbeitsbereich 6:** Integration neuer Funktionen in das Verbundfindmittel, die auf Grund der Ausdehnung auf mehr beteiligte Archive und wegen der neuen technischen und gesellschaftlichen Entwicklungen, besonders bei den Anforderungen an die Dienstleitungen der Archive, erforderlich werden

AP 6.1: Integration der Verlinkungsfähigkeit nach außen zu den Web-Angeboten der beteiligten Archive in das Verbundfindmittel. Es ist bereits jetzt möglich, URLs in den Findbuchtexten einzufügen. Diese Funktion sollte jedoch noch weiter unterstützt und bei der Eingabe der URL, z.B. mit einer „Durchsuchen“-Funktion und automatischer Integration der gewünschten Adresse ausgebaut werden.

AP 6.2: Erweiterung des Verbundmittels um eine Funktion für die Anzeige von Gruppen von Digitalisaten aus den Beständen, etwa ganzen Akten, neben der bereits möglichen Anzeige einzelner Bilder zur Illustration. Hierzu können die Ergebnisse des <daofind>-Projektes genutzt werden. Dort wurde der Editor für METS-Dateien, die die Metadaten für alle Images einer Akte bündeln und strukturieren, entwickelt. Zusammen damit wurde das Präsentationsmodell mit einer Orientierungsoberfläche über der Ebene zum Durchblättern der Akte entwickelt, auf der, wenn gewünscht, strukturelevante Schriftstücke etwa in einem scrollbaren Fenster zunächst mit dem oberen Drittel der Seite angezeigt werden. Eine Funktion für die Volltextsuche in den Digitalisaten ist in Vorbereitung und wird nach Fertigstellung ebenfalls für das Portal nutzbar sein.

Ergebnis aus Arbeitsbereich 6: Diese Entwicklungen geben den Archiven weitere bisher nicht als Angebot realisierte Möglichkeiten zur Gestaltung der Informationen im Verbundfindmittel. Sie können auf einer von ihnen selbst zu wählenden Ebene die Besucher des Portals auf ihr eigenes Internetangebot leiten oder hinweisen. Außerdem kann ihnen eine Funktion angeboten werden, mit der sie ganze digitalisierte Akten in einer dem Archivgut angepassten Präsentationsform anbieten können und die Images zusammen mit den Metadaten in das Verbundfindmittel einfügen können. Erforderliche Kapazität: 10 Personenmonate

### **Ergebnisse des Teilprojektes 2:**

In diesem Teilprojekt findet die erforderliche Softwareentwicklung statt. Hier wird die Suchmaschine auf den Stand gebracht, dass sie die erforderlichen Datenmengen reibungslos verwalten und die Suche mit der erforderlichen Performance durchführt. Außerdem werden die Mechanismen entwickelt, die den Archiven die Integration ihrer

Daten in das Portal erlauben und sie dabei weitest möglich unterstützen, damit diese Tätigkeiten ohne zusätzliche Ressourcen durchgeführt werden können. Schließlich findet die Öffnung des bisherigen Angebots um weitere, speziell für den dezentralen Betrieb des Portals erforderliche Funktionen statt. Außerdem wird das Verbundfindmittel mit den neusten und innovativen Funktionen versehen, die von den Besuchern des Portals, den potentiellen und geübten Archivbenutzern, in der nächsten Zeit verstärkt erwartet werden. Erforderliche Kapazität: 32 Personenmonate

### **Teilprojekt 3: Abschlussarbeiten und Vorbereitung für eine weitere Vergrößerung des Portals zu einem Archivportal D**

**Arbeitsbereich 7:** Aufbereitung der Projektdokumentation und Erstellung von Publikationen zu den im Projekt geleisteten Arbeiten

AP 7.1: Aufbereitung der Projektunterlagen für die Projektdokumentation und den Abschlussbericht.

AP 7.2: Abschließende Ergänzung der Webseite mit dem Abschlussbericht sowie abschließende Fertigstellung weiterer Texte, Folienpräsentationen, Screenshots, Linklisten und Manuskripte für Publikationen in Fachzeitschriften.

AP 7.3: Erstellung einer Dokumentation der Arbeitsweise des Portals für die beteiligten Archive und für die evtl. Werbung zukünftiger weiterer Partner.

Ergebnis aus Arbeitsbereich 7: Alle erforderlichen Texte und Illustrationen für die Öffentlichkeitsarbeit liegen vor und werden verteilt. Die Arbeiten finden zeitversetzt statt. Erforderliche Kapazität: 3 Personenmonate

**Arbeitsbereich 8:** Prüfung der Leistungsfähigkeit des erreichten Standes und eventueller weiterer Entwicklungsanforderungen für ein Archivportal D

AP 8.1: Tests mit künstlich massiv vergrößerten Mengen und Feststellung dabei evtl. auftretender Probleme sowie Überlegungen zu möglichen Lösungen. Dazu müssen wieder Testumgebungen geschaffen und Checklisten aufgestellt werden. Die Tests müssen dokumentiert und in einem Bericht zusammengefasst werden.

AP 8.2: Überblick über den Markt und zwischenzeitliche technologische Entwicklungen für das Portalmanagement bei einer Erhöhung von Umfang und Performance bei weiter deutlich steigenden Zahlen der beteiligte Archive und einbezogener Bestände. Es sollte hier festgestellt werden, ob es nach zwei Jahren Projektlaufzeit neue technologische Entwicklungen im Bereich der Verwaltung großer Portale gibt. Eine solche Entwicklung ist anzunehmen, da die Nutzung von Portalen ein sehr expandierender Geschäftsbereich ist. Es sollte festgestellt werden, ob neue Entwicklungen sinnvoll für eine Vereinfachung des Portalbetriebs genutzt werden können. Auch diese Untersuchung ist zu dokumentieren.

AP 8.3: Erstellung eines Berichts über die Prüfungen und vorgeschlagene Lösungsmöglichkeit mit Abwägung von Alternativen. Dieser Bericht soll möglichst weitgehend zeigen, was mit dem aktuellen Stand der Entwicklung möglich ist und wo eventuelle Grenzen absehbar sind. Daraus folgernd sollten Entwicklungserfordernisse für ein Archivportal D beschrieben werden.

Ergebnis des Arbeitsbereichs 8: Das Ergebnis soll ein Bericht sein, aus dem hervorgeht, wie leistungsfähig das ausgebaute Portal SED-Archivgut ist, und welche Erweiterungen erforderlich sind, um es zu einem Archivportal D auszubauen. Dieser Bericht soll als Grundlage für einen Förderantrag für den Aufbau eines deutschen Archivportals genommen werden können. Erforderliche Kapazität: 3 Personenmonate

### **Ergebnis des Teilprojektes 3:**

Hiermit wird das Projekt abgeschlossen und Rechenschaftsfähigkeit gegenüber den Geldgebern wie den Fachkollegen und der breiten Öffentlichkeit abgesichert. Außerdem wird die Verwendbarkeit der in diesem Projekt vorgenommenen Entwicklungen für darauf aufbauende weitere Perspektiven geklärt. Erforderliche Kapazität: 6 Personenmonate

## **4 Beantragte Mittel**

### **4.1 Personalbedarf**

Der oben dargestellte Personalbedarf ergibt sich aus einem deutlichen Schwerpunkt des Projektes in der fachbetonten Koordinierung und Entwicklung der im Arbeitsplan genannten Bereiche. Dazu kommt die Absprache mit den bisher am Portal SED-Archivgut beteiligten Archiven. Da innovative Softwareentwicklung vorgenommen werden muss, fällt viel Detail-Konzeptionierung an ebenso wie das Testen vorgelegter Entwicklungsleistungen. Dazu sollten Testszenarien entwickelt werden, damit möglichst zügig die Einsatzbereitschaft hergestellt werden kann.

Zur Projektbetreuung, Detaillierung, Steuerung und Kontrolle der Entwicklungsleistungen wird eine Mitarbeiterin oder ein Mitarbeiter mit archivischen Fachkenntnissen und IT-Kenntnissen nach BAT IIa für 2 Jahre benötigt (Person 1).

Außerdem wird eine Mitarbeiterin oder ein Mitarbeiter nach BAT IIa ebenfalls für 2 Jahre für die Vorbereitung einschließlich des Aufbaus der Testumgebung, für die konkrete Durchführung von Testarbeiten mit ihrer Dokumentation und der Berichterstellung und für die Aufbereitung von Content für eine Testinstallation und Pilotanwendung benötigt (Person 2).

Von den beiden Mitarbeiterinnen oder Mitarbeitern wird außerdem die detaillierte Abstimmung mit den Partnerarchiven sowie die Vorbereitung und Auswertung des geplanten Workshops erwartet. Dazu gehört auch die Erstellung der Mappings zwischen den Datenformaten vor Ort und dem Importformat in Kooperation mit den jeweiligen Archiven. Die Zuständigkeiten sollen schwerpunktmäßig wie folgt verteilt werden:

Person 1:

- AP 1.1 - 1.3                      2 PM
- AP 3.1 - 3.2 zu 50%            3 PM
- AP 4.1 - 4.2                      10 PM
- AP 6.1                              6 PM
- AP 7.1 - 7.3                      3 PM

Summe:                                24 PM

Person 2:

• AP 2.1 - 2.2	2 PM
• AP 3.1 - 3.2 zu 50%	3 PM
• AP 5.1 - 5.3	12 PM
• AP 6.2	4 PM
• AP 8.1 - 8.3	<u>3 PM</u>
Summe:	24 PM

Die Softwareentwicklung wird zum großen Teil außer Haus vergeben werden. Die Gründe dafür sind, dass sie sehr aktuelle Informatikkenntnisse erfordert und oft Absprachen in einem Team von Programmierern mit unterschiedlichen Spezialisierungen wie Java-Programmierung, Eclipse-Entwicklung, Lucene-Programmierung, XML- und XSLT-Erfahrungen nötig macht, wie es nur in einer Firma möglich ist. Dazu kommt, dass die im Portal SED-Archivgut bisher eingesetzte Lucene-Programmierung eine Eigenentwicklung einer Firma ist und von ihr lizenziert wird (von Startext in Bonn, Berlin und Tomsk).

Der Personalbedarf insgesamt:

**2 x BAT IIa für zwei Jahre**

#### **4.2 Wissenschaftliche Geräte**

Werden als Eigenleistung gestellt.

#### **4.3 Reisen**

Zur Information und Koordination der beteiligten Archive sollen kleinere Workshops und Reisen der Projektmitarbeiter zu den verschiedenen Einrichtungen erfolgen.

#### **4.4 Sonstige Kosten für Aufträge an Dritte**

Die Suchmaschine MidosaSEARCH ist in verschiedenen Ausbaustufen einsetzbar. In einer Light-Version mit Mengenbegrenzung auf 5 Bestände wird sie von der Archivschule verteilt und u.a. vom DHI in Rom eingesetzt. Für unbegrenzte Mengen an Beständen und Verzeichnungseinheiten wird sie von der Firma Startext pro eingesetztem Prozessor zum Preis von 9.500 € lizenziert. Die Firma hat die Kosten der Entwicklung selbst getragen. Das Bundesarchiv hat bisher drei Lizenzen erworben und wird die Weiterentwicklung für den Einsatz für seine eigenen Online-Präsentationen mit einer ersten umfangreichen Skalierung finanzieren. Für den Betrieb des Verbundfindmittels werden auf Dauer weitere vier Lizenzen à 9.500 € = 38.000 € erforderlich sein. Der Betrag ist abgewogen gegenüber dem Betrag für den vollständigen Erwerb der Anwendung mit allen dazu bisher angefallenen Entwicklungskosten in Höhe von 160.000 € oder die Erweiterung der drei im

Bundesarchiv eingesetzten Lizenzen zu einer Site-Lizenz ohne Begrenzung der Prozessorenzahl im Wert von 10 Lizenzen, also 95.000 €.

Für den Betrieb des erweiterten Portals wird ein stärkerer Server mit voraussichtlich vier Prozessoren benötigt, um die erforderliche Performance zu erreichen. Er wird vom Bundesarchiv im Rahmen seines Blade-Server-Centers bereitgestellt.

Die folgenden Entwicklungsaufträge sollen vergeben werden.

Die Firma Startext bietet die Erweiterungsarbeiten an und hat folgende Zeiten für die Softwareentwicklung berechnet:

## Pflichtenheft für die Softwareentwicklung

Inhalt:

- P 1: Erweiterung der Suchmaschine um die dritte Ebene
- P 2: Anpassung der Editoren
- P 3: Coaching bei der Erstellung der Mappings
- P 4: Ausbau von Kapazität und Performance der Suchmaschine
- P 5: Anpassung / Entwicklung eines Harvesters
- P 6: Erstellung eines WebDAV-Zugangs
- P 7: Unterstützung für die Einfügung von externen Links in die Suchmaschine
- P 8: Integration von METS für Images

### P 1 Erweiterung der Suchmaschine um die dritte Ebene

Entspricht Arbeitspaket AP 4.2 und Punkt 2 des Arbeitsplans

Für die Integration der Findbücher der teilnehmenden Archive in die übergreifende Suche muss neben den beiden Ebenen für die Auswahl der Archive und der jeweils von ihnen angebotenen Beständeübersichten eine dritte Ebene für die Findbücher in die Suchmaschine eingefügt werden. Das betrifft auch die Aggregation der Anzeige der Treffer. Sie werden in einer aufklappbaren Strukturansicht die der Tektonik entspricht, angeboten. Damit wird eine sinnvolle Reduktion der Anzahl angezeigter Treffer erreicht, ohne dass die Trefferzahl künstlich auf eine willkürlich angenommene Zahl beschränkt werden muss. Zur Begleitung der externen Softwareentwicklung werden intern regelmäßige Tests stattfinden, zu denen ein Testscenario aufgebaut und eine Checkliste erstellt werden.

MidosaSEARCH war ursprünglich für die findmittelübergreifende Suche über die Bestände eines Archivs konzipiert. Sie wurde vom Netzwerk SED-Archivgut bereits für Archive und Bestände eingesetzt. Dort fehlt deshalb im Moment die untere Ebene. Es solle jetzt gezielt die obere Ebene, nämlich die strukturierte Anzeige der Archive neu konzipiert und entwickelt werden, damit sie dem Bedarf genauer angepasst werden kann. Die neue, oberste Ebene soll aus einer regionalen und ggf. einer spartenbezogenen Gliederung bestehen, die zu den teilnehmenden Archiven führt.

#### Einzelaufstellung Erweiterung der Suchmaschine

##### 1.1 Revision der Verlinkung zwischen den Ebenen

Die bestehende Verlinkung zwischen Findmitteln und Beständeübersicht wird

daraufrhin durchgegangen und dokumentiert, an welchen Stellen die Erweiterung Änderungen in der Programmlogik erfordert, weil sie bisher die Eigenschaft von nur zwei Ebenen unterstützt. Soweit erforderlich, wird der Programmcode an den neuen Bedarf angepasst.

- 1.2 Aufstellung der Testszenarien  
Mitwirkung mit dem technischen Beitrag an der Erstellung der Testszenarien und der Aufstellung der Checkliste.
- 1.3 Aufbau der Gliederung der Archive nach Regionen und Sparten  
Die Pflege der Beschreibung der regionalen und spartenbezogenen Gliederung der Archive erfolgt im EAD Format. Zum Einsatz kommt der MEX Editor. Das Arbeitspaket umfasst die Erstellung des Definition Set für die Gliederung sowie die exemplarische Erfassung des Teils der Gliederung für die teilnehmenden Archive.
- 1.4 Verknüpfung der Gliederung der Archive mit den Informationen über die Einrichtungen  
Für die Gliederung der Anzeige der Archive soll EAD verwendet werden. Da für die Informationen zu den Archiven das Format EAG, das für eine dezentrale Pflege der Daten geeignet ist, eingesetzt werden soll, werden aus der Gliederung heraus Links zu den EAG-Dateien führen. Das Arbeitspaket beinhaltet die Anpassung der Indexierung für die getrennte Datenhaltung von Gliederung und EAG-Dateien. EAG wurde in Spanien für den Censo Guía entwickelt, in dem die Erschließungs-Informationen von 40.000 spanischsprachigen Archiven weltweit durchsucht werden können. Mit EAG werden dort die Angaben zu den Einrichtungen wie die Adresse und die Öffnungszeiten verwaltet.
- 1.5 Verknüpfung der EAD Beständeübersicht mit den Informationen zu den Einrichtungen  
Die EAG-Beschreibung eines Archivs verweist auf ihre Beständeübersicht in EAD. Für die Indexierung wird der Rückbezug von der Beständeübersicht zur EAG-Beschreibung benötigt. Das Arbeitspaket löst die Sonderbehandlung dieser hierarchischen Verknüpfung.
- 1.6 Anpassung der Verknüpfungen der Findbücher an die Einführung der 3. Ebene  
Die Suchmaschine stellt die Treffer zu Verzeichnungseinheiten im Kontext der Beständeübersicht dar. Das Arbeitspaket behandelt die Erweiterung des Kontextes um die Gliederung der dritten Ebene. Eine Besonderheit liegt in der Unterdrückung redundanter Information, die sich aus Wiederholungen im Wurzelknoten der Beständeübersicht und EAG-Datensatz ergeben kann und die in der Darstellung des Kontextes störend wirkt.
- 1.7 Anpassung der Listen bei der Trefferanzeige zur Darstellung des Kontextes auf 3 Ebenen  
Die Trefferliste zeigt in der bestehenden Version eine zweistufige Darstellung mit Bestandsnamen und Klassifikationspunkt im Findmittel. Das Arbeitspaket beinhaltet die Entwicklung einer dreistufigen Darstellung, die den Kontext des teilnehmenden Archivs umfasst. Die Entwicklung ist prototypisch mit mehreren Zyklen vorgesehen, um den unterschiedlichen Benutzungsfällen gerecht zu werden: betrifft große Treffermengen in einem Archiv gegenüber vielen Archiven in der Treffermenge, die jeweils nur wenige Treffer enthalten. Diese Entwicklung baut auf der Erweiterung für das Bundesarchiv auf, die im August 2006 begonnen hat und eine Anzeige der jeweils übergeordneten Struktur neben der zweistufigen Trefferliste vorsieht.
- 1.8 Anpassung der Möglichkeit einer pauschalen Eingrenzung der Recherche für die 3. Ebene für die Begrenzung der Recherche auf Gruppen von ausgewählten Archiven

Das Bundesarchiv erbringt in Eigenleistung die Erweiterung der Suchmaschine um die Möglichkeit, mehrere Gliederungspunkte der Tektonik in der Beständeübersicht pauschal zur Eingrenzung der Recherche zu verwenden, statt alle darunter liegenden Findbücher einzeln auswählen zu müssen. Diese Entwicklung wird die Funktion auf die dritte Ebene ausweiten und dient der Eingrenzung der Recherche unter Verwendung der Gliederung der dritten Ebene.

## P 2 Anpassung der Editoren

Entspricht Arbeitspaket AP 5.1 und Punkt 3 des Arbeitsplans

Für eine gemeinsame zentrale Anzeige der Erschließungsdaten und die darin anzubietende übergreifende Recherche ist ein einheitliches Datenformat erforderlich. Mit EAD liegt ein international breit eingesetztes Format vor, das in verschiedenen Anwendungen in Deutschland, u.a. beim BAM-Portal, genutzt wird und in der Lage ist, die für die zentrale Suche erforderlichen Findbuchdaten in hierarchischer Gliederung bis zur untersten Ebene darzustellen. Eingesetzt als Arbeitsformat des Portals ermöglicht es die Weitergabe der Informationen aus dem Verbundfindmittel an internationale Gateways.

Mit Hilfe der im daofind-Projekt des Bundesarchivs prototypisch entwickelten Editoren auf Eclipse-Basis können den Archiven lizenzfreie Open-source-Werkzeuge zur Konversion ihrer eigenen Daten in das zentrale Format bereitgestellt werden. Die Editoren werden dazu an das jeweils eigene Datenformat angepasst und erlauben entweder die automatische Konversion im Batch-Mode oder eine evtl. gewünschte Überarbeitung etwa zur Einfügung von Links oder Images.

Den Archiven soll die Arbeit bei der Aufbereitung eigener Daten für die Integration in das Portal durch die an ihre jeweiligen Anforderungen angepassten Werkzeuge erleichtert werden. Die Anpassung soll mit ihnen zusammen vorgenommen werden. Dazu sind von den Projektmitarbeitern Tests mit Daten aus den jeweiligen Archiven durchzuführen. Zu diesen Entwicklungen gehört die Anpassung des EAD-Editors für Findbücher und Beständeübersichten und des EAC-Editors für die Angaben zu den Herkunftsstellen, die von der Beständeübersicht aus aufgerufen werden können.

Werkzeuge für die Konversion von Daten aus den Basys-Anwendungen des Bundesarchiv nach EAD liegen vor. Die Feinabstimmung für das zentrale EAD-Format findet in Eigenleistung des Bundesarchivs statt.

Neu programmiert werden muss der Editor für EAG-Daten, der die Informationen zu den Archiven in einem mit den Plänen für das europäische Portal kompatiblen Format erfassen und bearbeiten läßt. Mit seiner Hilfe sollen die Archive auch die Grundinformationen zu ihrer Einrichtung wie Adresse, Öffnungszeiten etc. erfassen und pflegen können. Dazu gehört auch die Umstellung der Suchmaschine Midosasearch von den bisher noch verwendeten deutschsprachigen Tags der DTD Midosaxml, die weitgehend Übersetzungen der englischen Tags von EAD sind, auf das im Herbst 2006 zu erwartende Schema von EAD.

## Einzelaufstellung für die beispielhafte Entwicklung von Konversionstools

### 2.1 Beispiel 1: Augias

#### 2.1.1 Augias – Datenübernahme

Das Archivprogramm Augias strukturiert Erschließungsangaben nach ISAD-G und bietet mehrstufige Erschließung. Die Daten werden dabei in der aktuellen Version 8 mit MS SQL Server 2000 oder Oracle 8.x/10g gespeichert. (vgl. Produktinformation unter [www.augias.de](http://www.augias.de)) Eine Speicherung unter MySQL (Linux) ist laut Hersteller in Vorbereitung. Das Programm bietet eine Exportfunktion, die den gewünschten Datensatz in eine MS-Access Datenbank überführt. Das Exportformat unterscheidet sich in den Tabellennamen, die Inhalte sind ansonsten deckungsgleich.

Das Arbeitspaket enthält die Erstellung eines Programms, welches die Daten aus der nativen Version 8 Datenbank oder der Access Datenbank in das hierarchische EAD XML Dokument mit der Ausprägung für das Portal überführt.

#### 2.1.2 Klassifikationskonversion

Die Übernahme der Klassifikationsstruktur von Augias Daten ist aufwendig: Es gilt, die Hierarchie der Klassifikation aus den Klassifikationstiteln abzuleiten, die als Dezimalklassifikation in den Text des Feldes eingefügt ist. Eine technisch konsistente Verlinkung der Hierarchie ist in den Daten nicht enthalten. In überlassenen Daten sind startext Fälle bekannt, in denen die Dezimalklassifikation überschrieben ist. Weiter sind Datensätze in der Klassifikation enthalten, die als „Dummies“ keine Verwendung für die Konversion finden.

#### 2.1.3 Feinabstimmung Vorschau Augias

Die Überprüfung des Konversionsergebnisses wird an Hand einer Vorschau (Preview) der EAD Präsentation des Konversionsergebnisses durchgeführt. Der Bearbeiter gewinnt einen schnellen Überblick zum Konversionsergebnis. Im Arbeitsfluss bietet die Vorschau einen guten Kompromiss, um im weitgehend automatisierten Arbeitsablauf eine visuelle Qualitätskontrolle einzubetten.

### 2.2 Beispiel 2: Austauschformat des Projektes Retrokonversion (SAFT)

#### 2.2.1 Retrokonversion Sachakten und Urkunden

Das XML Austauschformat des Retrokonversionsprojekts ist Ausgangsbasis, um per XSLT Script die Findmittel in das hierarchische EAD-XML-Dokument entsprechend dem gemeinsam genutzten Profil zu konvertieren. Dazu ist das vorliegende Mapping, das mit Unterstützung des Bundesarchiv angefertigt wurde, an die Verwendung im Verbundfindmittel anzupassen.

#### 2.2.2 Klassifikationskonversion

Die Klassifikation ist im XML Austauschformat (SAFT) des DFG Projekts Retrokonversion enthalten. Das Arbeitspaket besteht aus der Erstellung eines XSLT Scripts zur Konversion der Klassifikation in das EAD Format. Das Austauschformat ist überaus flexibel. Mit dem geschätzten Zeitaufwand wird die exemplarische Konversion der Klassifikation realisiert, die von teilnehmenden Archiven nach eigenem Ermessen angepasst werden wird.

#### 2.2.3 Feinabstimmung Vorschau Austauschformat

Die Überprüfung des Konversionsergebnisses wird an Hand einer Vorschau (Preview) der EAD Präsentation des Konversionsergebnisses durchgeführt. Der Bearbeiter gewinnt einen schnellen Überblick zum Konversionsergebnis. Im Arbeitsfluss bietet die Vorschau einen guten Kompromiss, im weitgehend automatisierten Arbeitsablauf eine visuelle Qualitätskontrolle einzubetten.

- 2.3 Beispiel 3: MidosaXML
  - 2.3.1 MidosaXML Datenübernahme
 

Das Arbeitspaket baut auf der Nachnutzung von Scripten auf, die vom Bundesarchiv als Eigenleistung erstellt werden. Die Arbeiten selbst bestehen in der Anpassung der Scripte an das EAD Format in der Ausprägung für das Portal.
  - 2.3.2 Feinabstimmung Vorschau MidosaXML
 

Die Überprüfung des Konversionsergebnisses wird an Hand einer Vorschau (Preview) der EAD Präsentation des Konversionsergebnisses durchgeführt. Der Bearbeiter gewinnt einen schnellen Überblick zum Konversionsergebnis. Im Arbeitsfluss bietet die Vorschau einen guten Kompromiss, im weitgehend automatisierten Arbeitsablauf eine visuelle Qualitätskontrolle einzubetten.
  
- 2.4 Erstellung eines dezentral zu nutzenden Editors für EAG-Daten zu den Einrichtungen
  - 2.4.1 Erstellung des Definitions Sets auf der Grundlage des EAG-Schemas
 

Die vorhandene XML Schema Definition (XSD) wird in einen Mex Definition Set überführt und ein einheitliches Nutzungsprofil wird festgelegt.
  - 2.4.2 Erstellung von Formularen für den formularbasierten Zugriff
 

Die Darstellung der EAG Inhalte wird in Formularen definiert. Die Formulare bieten einen alternativen Zugriff auf die EAG Daten. Der Benutzer kann dann zwischen Baumansicht und Formularansicht wechseln.
  - 2.4.3 Formalprüfung
 

Dieses Arbeitspaket besteht aus Überprüfungen auf die inhaltliche Richtigkeit von EAG Daten, die innerhalb des EAG Definition Set nicht abgelegt werden können. Beispiele sind die Überprüfung der Postleitzahlen oder die Prüfung der Öffnungszeiten auf plausible Werte.
  
- 2.5 Anpassung des EAC-Editors für die dezentrale Nutzung und Erfassung von Daten für die Beständeübersichten
  - 2.5.1 Anpassung EAC Definition Set
 

Dieses Arbeitspaket besteht aus den Anpassungen am EAC Definition Set für den Mex Editor. Das bestehende Definition Set für EAC wird um die fehlenden Bestandteile für die Nutzung von EAC im Rahmen des Portals erweitert. Durch die Kombination von EAC und EAG wird durch eine Realisierung mit unterschiedlichen Namespaces innerhalb des EAC Definition Set ermöglicht.
  - 2.5.2 Erstellung von Formularen für formularbasierten Zugriff
 

Die Darstellung der EAC Inhalte wird in Formularen definiert. Die Formulare bieten einen alternativen Zugriff auf die EAC Daten. Der Benutzer kann dann zwischen strukturierter Baumansicht und Formularansicht wechseln.
  
- 2.6 Anpassung der Validierungsfunktionen
  - 2.6.1 Validierung von EAD
 

Durch den Einsatz des Definition Set für EAD Dateien im Mex Editor muss sichergestellt sein, dass eine bearbeitete EAD Datei aus wohlgeformtem XML besteht und nach den Regeln der allgemeinen EAD Definition (entsprechend dem im Herbst erscheinenden Schema) gültig ist, sowie dass sie der für das Verbundfindmittel verwendeten Definition von EAD entspricht, so dass eventuelle Abweichungen davon angezeigt werden.
  - 2.6.2 Validierung EAC
 

Durch den Einsatz des Definition Set für EAC Dateien im Mex Editor muss sichergestellt sein, dass eine bearbeitete EAC-Datei aus wohlgeformtem XML

besteht und nach den Regeln der allgemeinen EAC Definition (XSD) gültig ist, sowie dass sie dem für das Verbundfindmittel verwendeten Profil von EAC entspricht, so dass eventuelle Abweichungen davon angezeigt werden.

#### 2.6.3 Validierung EAG

Durch den Einsatz des Definition Set für EAG Dateien im Mex Editor muss sichergestellt sein, dass eine bearbeitete EAG Datei aus wohlgeformtem XML besteht und nach den Regeln der allgemeinen EAG Definition (XSD) gültig ist, sowie dass sie dem für das Verbundfindmittel verwendeten Profil von EAG entspricht, so dass eventuelle Abweichungen davon angezeigt werden.

#### 2.6.4 Konsistenzprüfung METS

Durch den Einsatz des Definition Set für METS Dateien im Mex Editor muss sichergestellt sein, dass eine bearbeitete METS-Datei aus wohlgeformtem XML besteht, den Regel der Definition im Schema für METS entspricht und valide ist und dem für das Portal definiertem Profil entspricht. Eventuelle Abweichungen sollen angezeigt werden, damit sie direkt korrigierbar sind. Dieses Arbeitspaket umfasst zudem eine Überprüfung der Konsistenzen von Verknüpfungen auf externe Dateien (File Pointer), die für die Verknüpfung der METS-Datei mit anderen XML-Files und den erfassten Digitalisaten wichtig sind.

#### 2.7 Erstellung eines Stylesheets für die Beständeübersichten

Das Stylesheet für die Darstellung der Beständeübersicht in der Kombination von EAD (für die Tektonik und Beschreibungen der Bestände) und EAC (für weitere Informationen zu den Herkunftsstellen) auf der Grundlage eines mit den beteiligten Archiven entwickelten Konzepts kann eine integrierte Darstellung mit allen oder ausgewählten Informationen oder die Aufteilung der Daten in verschiedene Ansichten bieten. Das genaue Konzept sollte sich anlehnen an das in MidosaSEARCH vorhandene und im Projekt daofind weiterentwickelte Präsentationsmodell.

#### 2.8 Erstellung der Stylesheets für die Findbücher

Die Stylesheets für die Darstellung der Findbücher sollen die vorhandene Präsentation an die verwendete EAD-Kodierung anpassen und die Links zu Digitalisaten verwalten.

#### 2.9 Anpassung des Präsentationsmodells für Digitalisate

Die Präsentation der Digitalisate wird über die Structural Map in METS gesteuert. Diese Funktionen sollen an den speziellen Bedarf des Portals angepasst werden. Außerdem sollen die Darstellungen der Digitalisate um weitere Kontext-Informationen und Links zum Findbuch ergänzt werden, so dass auch der Weg zurück über die Abbildung zum Findbuch möglich ist.

## P 3 Coaching bei der Erstellung der Mappings

Entspricht Arbeitspaket AP 5.1 und Punkt 3 des Arbeitsplans

Den Archiven soll die Arbeit bei der Aufbereitung eigener Daten für die Integration in das Portal durch angepasste Werkzeuge erleichtert werden. Deshalb gehört zu den Erweiterungen des Portals der Aufbau eines Werkzeugkastens für die beteiligten Archive, mit dem sie ihre Erschließungsdaten aus eigenen Datenformaten, soweit gewünscht und erforderlich, in das gemeinsame Austauschformat bringen können. Voraussetzung ist die Erstellung von Mappings oder Cross-Walks zwischen den ursprünglichen Formaten und dem Zielformat. Die Mappings werden mit den beteiligten Archiven an Hand ihrer intern genutzten Datenstrukturen erstellt. Dabei werden sich viele Arbeiten wiederholen. Außerdem kann auf

vorhandene Mappings wie MidosaXML <-> EAD und Austauschformat (NRW) <-> EAD zurückgegriffen werden. Der größte Teil der Arbeit kann deshalb von Projektmitarbeitern erledigt werden. Sie benötigen jedoch Unterstützung für die technischen Fragen.

### Einzelaufstellung zum Coaching

- 3.1 Unterstützung bei Erstellung der gemeinsam zu nutzenden Profile von EAD, EAC, EAG und METS für das Verbundfindmittel.  
EAD ist ein sehr flexibler Standard, der viele Möglichkeiten der Nutzung bietet. Anderserseits ist er relativ komplex, insofern als eine Entscheidung für die Nutzung eines Elementes oder einer Elemente-Elemente- oder einer Elemente-Attribute-Kombination Konsequenzen für die Nutzung anderer Elemente hat. Die für das Verbundfindmittel zu erstellenden Profile ("Best Practice Guidelines") sollten dabei die bisherige Verwendung vor allem von EAD an den Stellen, an denen EAD bereits, etwa für den internationalen Austausch genutzt wird, zu Grunde legen. Das Bundesarchiv hat bereits umfangreiche Vorarbeiten für die Nutzung von EAD für den internationalen Austausch und als Import- und Exportformat zwischen der Datenbank Basy und MidosaXML sowie beim Mapping des Austauschformats der Retrodigitalisierungsprojektes (SAFT) geleistet und ist in die Beratungen zur Nutzung von EAD im BAM-Projekt einbezogen. Es sind deshalb umfangreiche hausinterne Kenntnisse vorhanden. Sie sollten durch die technischen Fachkenntnisse der Firma ergänzt werden.
- 3.2 Unterstützung bei der Aufstellung der Mappings für die Übertragung der Daten aus den dezentralen Datenstrukturen in das gemeinsame Profil  
Die Erstellung der Mappings für die Festlegung, welche Ursprungsdaten aus den jeweiligen Archivsystemen in welche EAD-Elemente oder Element-Attribut-Kombinationen übertragen werden sollen, muss viele Details berücksichtigen. Sie erfordert zum einen archivarische Kenntnisse, um die jeweilige Art der Nutzung schnell und genau zu verstehen. Deshalb wird es eine wichtige Aufgabe der Projektmitarbeiter sein. Zum anderen erfordert sie Kenntnisse der XML-Technik und der Erstellung von Transformations-Stylesheets. Deshalb sollen regelmäßige Besprechungen der Projektmitarbeiter mit Mitarbeitern der Firma stattfinden, in denen Detailprobleme besprochen und gelöst werden. Die Besprechungen sollen im Umfang von jeweils 2 Tagen innerhalb von 8 Monaten stattfinden.

## P 4 Ausbau von Kapazität und Performance der Suchmaschine

Entspricht Arbeitspaket AP 4.1 und Punkt 1 des Arbeitsplans

Gegenüber der im Bundesarchiv vorgenommenen Erweiterung auf eine Datenmenge von 1,5 Mill. Verzeichnungseinheiten soll die Suchmaschine für das Verbundfindmittel mit Orientierung auf die Integration unterschiedlicher Archive, verschiedenartiger Bestände und unterschiedlicher Ausgangsformate von Erschließungsdaten zunächst auf eine Kapazität von etwa 4 Mill. Verzeichnungseinheiten gebracht werden. Dabei sind regelmäßige Tests erforderlich. Dazu muss ein Testszenario aufgebaut und eine Checkliste erstellt werden.

### Einzelaufstellung zum Ausbau der Suchmaschine

- 4.1 Ausbau der Suchmaschine  
Die Suchmaschine soll auf die parallele Nutzung von 4 Prozessoren eingestellt werden, die versetzt arbeiten und dadurch die Ausführung von Suchanfragen deutlich beschleunigen.

#### 4.2 Beschleunigung der Anzeige

Die Performance ist neben der Durchsuchung der Datenbestände selbst von der Lieferung und der Darstellung der Ergebnisse abhängig. Durch weitere Komprimierung und Aufteilung der zu übermittelnden Daten in leicht übertragbare Pakete wird die Geschwindigkeit der Anzeige noch einmal deutlich erhöht.

#### 4.3 Anpassung der Präsentationsstylesheets

Die Ergebnisse der Suche werden zur Navigation von Fundstelle zu Fundstelle im Kontext angeboten. Bei größeren Gliederungsgruppen soll die Möglichkeit gegeben sein, mit Aufwärts- und Abwärtspfeilen weitere Portionen ohne Zeitverzögerung zu laden.

## P 5 Anpassung / Entwicklung eines Harvester

Entspricht Arbeitspaket 5.2 und Punkt 4 des Arbeitsplans

Für die Abholung bereit gestellter dezentralen Daten und ihre Integration in das Verbundfindmittel wird ein Harvester benötigt, der evtl. aus anderen Projekten übernommen und angepasst werden kann, und mit dem die dezentral bereitgestellten Daten vor Ort indexiert und im erforderlichen Umfang kopiert werden. Dabei soll eine Prüfung stattfinden, ob ein vorhandener Harvester übernommen und angepasst werden kann.

Dazu muss zunächst festgestellt werden, welche vergleichbaren Harvester bisher entwickelt wurden und zur Verfügung gestellt werden könnten. Dann ist der erforderliche Anpassungsaufwand abzuschätzen und mit dem Aufwand einer Neuentwicklung sowie den Kosten für einen dauerhaften Einsatz und später notwendige Updates abzuwägen.

### Einzelaufstellung Harvester

#### 5.1 Evaluierung vorhandener Open Source Harvester

Ziel ist die Integration einer Open Source Komponente. Anforderungen sind: Ablauffähigkeit unter Linux, Integrierbarkeit in LDAP, Authentifikationsmechanismen, Harvesting auch über gesicherte Wege mit Authentifikation, Implementation bevorzugt in JAVA, um die Vielfalt verwendeter Sprachen und damit zukünftigen Wartungsaufwand gering zu halten.

Zwei Open Source Komponenten werden daraufhin evaluiert, wieweit die funktionalen Anforderungen erfüllt werden. Die Systeme werden in einem Testenvironment installiert.

Zum Zeitpunkt der Antragsstellung sind folgende Systeme vorausgewählt:

OCLC OAI Harvester, Referenz CDLIB

PKP Harvester 2, Referenz Netbib

Dazu soll, wenn möglich, ebenfalls der im Ariadne Projekt in Greifswald entwickelte Harvester in die Prüfung einbezogen werden.

#### 5.2 Anpassung des ausgewählten Harvesters für das Verbundfindmittel

Die Arbeiten bestehen in der Inbetriebnahme des Harvesters und umfassen die Parametrisierung sowie den Test mit ausgewählten Pilotteilnehmern.

Der Harvester arbeitet in 4 unterschiedlichen Profilen, die im EAG-Dokument von den bereitstellenden Archiven festgelegt werden:

Findbücher ohne Digitalisate in METS, jedoch Verifikation der Links auf METS und Digitalisate

Findbücher mit METS, aber noch ohne Digitalisate, Verifikation der Links auf Digitalisate

Findbücher, METS und Digitalisaten

Beständeübersicht

Aktualisierungen erfolgen für Findbücher und Beständeübersichten an Hand von periodisch abgefragten Zeitmarken.

#### 5.3 WEB Anwendung für den Zugang zum Harvester

Die Arbeiten bestehen in der Realisierung bzw. Anpassung einer WEB-Schnittstelle zur Steuerung des Harvesters. Die Steuerungsfunktionen sind „Start“, „Stop“, „Status“ und „Reload“ für einzelne Archive. Der Zugang erfolgt passwordgeschützt.

#### 5.4 EAG Extension für den Zugang zum Harvester

Die teilnehmenden Archive wählen im EAG Dokument die Form ihrer Belieferung:

- Harvester
- WebDAV
- Physikalische Übermittlung auf Datenträger

Im Fall der Belieferung mit dem Harvester benennt das teilnehmende Archiv das Profil und beschreibt den Zugang des Harvesters zu den Daten. Hinzu kommen Parameter für die Transformation der URL's, um bei Bedarf die internen Links für den Zugang zu den Daten aufzubauen. Das Arbeitspaket sieht vor, den Definitionset im EAG-Dokument zu erweitern und den Harvester zu testen.

#### 5.5 Einweisung, Schulung und Einführungsunterstützung über 6 Monate

Das Arbeitspaket enthält die anfängliche Einweisung in die Funktionsweise und Bedienung sowie eine laufende Unterstützung während der ersten 6 Monate im Umfang von etwa 1 Tag pro Monat.

## P 6 Erstellung eines WebDAV-Zugangs

Entspricht Arbeitspaket AP 5.3 und Punkt 5 des Arbeitsplans

Für die Archive, die ihre Daten nicht dezentral bereitstellen können oder wollen, soll ein WebDAV-Zugang auf dem Server des Bundesarchiv eingerichtet werden, auf dem die Archive ihre eigenen Daten so bearbeiten, speichern und freigeben können wie auf den eigenen Speichern. Dazu soll ein browserbasierter Zugang angeboten werden.

Dazu muß der WebDAV-Zugang zu dem Server des Bundesarchivs einschließlich einer browsergestützten Nutzungsoberfläche für die Ablage und die Pflege der in das Portal einzubeziehenden oder bereits einbezogenen Daten erstellt und angepasst werden. Erfahrungen mit WebDAV-Arbeit liegen im Bundesarchiv bereits vor, da MidosaXML WebDAV-fähig ist, und die gemeinsame Arbeit an Findbüchern im internen Netz diese Funktion nutzt. Außerdem wird bereits seit einiger Zeit erfolgreich eine browsergestützte Eingabe und Pflege von Daten sowohl bei der zentralen Nachlassdatenbank (ZND) sowie bei dem Notfallregister für Archive NORA eingesetzt.

### Einzelaufstellung Entwicklung WebDAV Zugang

#### 6.1 Einrichtung WebDAV Server

Die Arbeiten bestehen aus der Einrichtung eines WEBDAV Servers. Voraussichtlich wird Apache oder Tomcat eingesetzt, wobei zur Authentifizierung LDAP verwendet wird.

#### 6.2 Client für die EAG Auswertung

Die teilnehmenden Archive erhalten vom Betreiber des Portals eine ID zugewiesen, die gleichzeitig die Wurzel im Dateisystem für die bereitgestellten Dokumente darstellt. Das teilnehmende Archiv trägt die zugewiesene ID im EAG Dokument ein. Das Arbeitspaket besteht aus der Erstellung einer WEB – Anwendung, alternativ einem Mail Client zum internen Gebrauch beim Portalbetreiber. Die Software übernimmt aus dem EAG Datensatz die Archiv-ID und richtet das Archiv als Benutzer unter LDAP ein.

#### 6.3 LDAP Konsistenz und Update mit EAG

Die EAG Daten sind Grundlage zur Identifikation der teilnehmenden Archive. Das Arbeitspaket dient dazu, technische Hilfen zur Sicherung der Konsistenz und zur Verhinderung ungewollter Änderungen bereitzustellen. Falls etwa EAG Dokumente zwischen Archiven als Beispiel ausgetauscht werden, können durch die evtl. nicht erneuerte ID Konsistenzprobleme auftreten. Verändert ein Archiv das übernommene Beispiel mit eigenen Daten und behält die ID des übernommenen Datensatzes bei, wird anschließend dieses EAG Dokument automatisch ausgewertet und die ID als Grundlage herangezogen, dann soll eine Sicherheitsmaßnahme greifen, damit die Daten nicht überschrieben werden.

#### 6.4 Workflow für das Update von Daemon für Neuzugänge

Stellt ein Archiv neue Findbücher per WebDAV bereit, dann sollen die Findbücher veröffentlicht und indexiert werden. Die Archive können sich als "ungeübte Nutzer" und "geübte Nutzer" einstufen. Diese Kategorien können jederzeit geändert werden, etwa bei Einarbeitung von neuem Personal.

Das Arbeitspaket dient dazu, einen Workflow zur Veröffentlichung zu implementieren, der für "ungeübte Nutzer" eine Begleitung vorsieht. Die "geübten Nutzer" erhalten eine automatisierte Benachrichtigung für den Status im Veröffentlichungsprozess: „Daten übernommen und formal geprüft“, „Verlinkung auf Konsistenz überprüft“, „Indexiert“, im Fall eines Update die Gegenüberstellung der Anzahl Verzeichnungseinheiten sowie eine Statistik zur Indexierung.

## P 7 Unterstützung für die Einfügung von externen Links in die Suchmaschine

Entspricht Arbeitspaket AP 6.1 und Punkt 6 des Arbeitsplans

Bisher wird von der Einstiegsseite des Portals aus ein direkter Link zu den Archiven angeboten. In Zukunft soll aus der übergreifenden Suche in den Findmittelinformationen auf beliebiger, von den Archiven wählbarer Ebene, auf deren eigene Präsentation verlinkt werden können. Dazu muss die Verlinkungsfähigkeit nach außen zu den Web-Angeboten der beteiligten Archive in das Verbundfindmittel leichter einsetzbar sein. Es ist bereits jetzt möglich, URLs in den Findbuchtexten einzufügen. Diese Funktion sollte jedoch noch weiter unterstützt und bei der Eingabe der URL, z.B. mit einer „Durchsuchen“-Funktion und automatischer Integration der gewünschten Adresse ausgebaut werden.

### Einzelauflistung zur Einfügung von Links in das Verbundfindmittel

#### 7.1 Assistent zur Erstellung von Links im HTML Code

Das Arbeitspaket besteht aus einer Erweiterung des MEX-Editors, um Links in den generierten HTML Code einzufügen. Der Editor erhält für die Initialisierung des Assistenten eine Drop-Funktion, mit der ein ausgewählter

Browser (z.B. Microsoft Explorer oder Mozilla) als Source der Verlinkung notiert wird.

Die laufende Bearbeitung erfolgt, indem mit dem angekoppelten Browser die zu verlinkende Seite aufgesucht wird. Mit dem Aktivieren des Editors übernimmt der Assistent den Link aus dem Browser und fügt den Link in die Zwischenablage ein. Dies erfolgt für XML als XLINK und für HTML als HREF. Bei der Eingabe wird der Cursor an die Stelle positioniert, an der der Link eingefügt werden soll.

#### 7.2 Protokoll eingefügter Links

Das Arbeitspaket besteht aus der Erweiterung des Editors, um die eingefügten Links als Protokoll aufzuzeichnen. Das Protokoll wird im XML Format erstellt, das neben dem Link eine vorhergehende IDREF sowie eine vorhergehende Signatur aufsucht und die Position zum eingefügten Link als XPATH Ausdruck darstellt.

## P 8 Integration von METS für Images

Entspricht Arbeitspaket AP 6.2 und Punkt 7 des Arbeitsplans

Das Verbundfindmittel sollte angesichts der begonnenen Diskussion um die Bereitstellung von Digitalisaten aus den Beständen der Archive um eine Funktion für die Präsentation von Images innerhalb der Findbücher erweitert werden. Dazu liegen ebenfalls Vorarbeiten aus dem Daofind-Projekt des Bundesarchivs vor.

Sie umfassen ein Werkzeug für die Integration von Images aus den Beständen in Online-Findmitteln und eine speziell für Schriftgut entwickelte Präsentationsform, bei der die Abbildungen zu je einer Verzeichnungseinheit in einem digitalen archivischen Objekt zusammengefasst werden und bei der Ansicht im Internet eine Orientierungsoberfläche und eine Blätteransicht miteinander verbunden sind. Dazu ist die Erweiterung des Verbundmittels um eine Funktion für die Anzeige von Gruppen von Digitalisaten aus den Beständen, etwa ganzen Akten, neben der bereits möglichen Anzeige einzelner Bilder zur Illustration zu erstellen. Der Editor nutzt für die Erfassung und Verwaltung der Metadaten zu den Digitalisaten den internationalen Standard METS, der alle Images einer Akte zu einem digitalen archivischen Objekt bündeln und strukturieren kann.

Bei der Integration von METS geht es um die Ausweitung möglicher Bildformate einschließlich einer internen hierarchischen Strukturierung etwa bei TIFF und PDF-Dateien. Dafür wird der Editor an verschiedene MIME-Types angepasst. Die Anpassung an den MIME-Type PNG und an JPEG sind bereits im Projekt daofind realisiert worden. PNG wird in der Pilotanwendung im Projekt daofind im Bundesarchiv für Images von Schriftgut eingesetzt.

Eine Funktion für die Volltextsuche in den Digitalisaten ist für das Bundesarchiv in Vorbereitung und wird nach Fertigstellung ebenfalls für das Portal nutzbar sein.

### Einzelaufstellung METS für das erweiterte Portal

#### 8.1 Erweiterung METS Definition Set

Diese Arbeiten umfassen Erweiterungen im METS Definition Set, der für den Prototypen des daofind-Projektes des Bundesarchivs entwickelt wurde. Er soll ergänzt werden um neue Funktionen für die programmgestützte und noch weiter automatisierte Erstellung von METS-Dateien.

#### 8.2 Transformation der Links

In diesem Paket wird die Verknüpfung und Darstellung von dezentralen

Dokumenten unterschiedlicher Archive an einer zentralen Stelle realisiert. Durch die Verbindung mit Informationen aus dem EAG Datensatz eines Archivs können relative URLs in extern verwertbare URL transformiert werden und die dezentral erstellten und bereitgestellten Dokumente zentral präsentiert werden.

### 8.3 Hierarchie für Multipage Digitalisate

Dieses Arbeitspaket besteht aus einer Erweiterung bei der programmgestützten Erstellung von METS Dateien. Die Hierarchie eines Multipage TIFFs oder einer PDF Datei wird beim Einlesen in eine METS Datei abgebildet. Dafür gilt es, die Grundlagen im Mex Definition Set und in den Einlesemechanismen des Mex Editors zu schaffen.

### 8.4 Mimetype Multipage TIFF

In diesem Arbeitsschritt wird die programmgesteuerte Erstellung von METS file pointern um die Möglichkeit des Verweises auf Multipage TIFF Dateien erweitert. Ebenso wird die Anzeige von Digitalisaten im Mex Editor um die Anzeige von TIFF Dateien erweitert.

### 8.5 Mimetype PDF

In diesem Arbeitsschritt wird die programmgesteuerte Erstellung von METS file pointern um die Möglichkeit des Verweises auf Multipage PDF Dateien erweitert. Hinzu kommt die Möglichkeit, das PDF Inhaltsverzeichnis unmittelbar in METS zu übernehmen.

### 8.6 Mimetype ELAK (Elektronischer Akt)

Um die Integration originär digitalen Archivgutes in die Findbücher zu ermöglichen, wird beispielhaft das XML-gestützte ELAK-Austauschformat integriert. In diesem Arbeitsschritt wird die programmgesteuerte Erstellung von METS file pointern um die Möglichkeit des Verweises auf ELAK Dateien erweitert. Es handelt sich um eine vorbereitende Arbeit, die eine entsprechende Schnittstelle aufbaut und offenhält. Das ELAK-Format wird etwa von Fabasoft eingesetzt, das von der KBSt für die öffentliche Verwaltung in Deutschland zertifiziert ist und zunehmend in Deutschland eingesetzt wird.

### 8.7 Mimetype RAW OCR

In Eigenleistung ermöglicht das Bundesarchiv, eine Volltextsuche in den Digitalisaten zu nutzen, die seitengenau über Text oder RAW OCR den Volltext positioniert. In diesem Arbeitsschritt wird die Möglichkeit zur programmgesteuerten und automatischen Erstellung von METS-file-pointern auf RAW OCR Dateien erweitert. Die Anzeige der Texte aus dem RAW OCR ist nicht vorgesehen. Sie dienen ausschließlich der Positionierung der Treffer und der Auswahl des anzuzeigenden Image.

## **5 Eigenleistungen für die Durchführung des Vorhabens, sonstige Voraussetzungen**

### **5.1 Zusammensetzung der Arbeitsgruppe**

Die Projektleitung wird von der Vizepräsidentin Frau Prof. Dr. Angelika Menne-Haritz übernommen. Nach Bedarf werden weitere Führungskräfte und Mitarbeiterinnen oder Mitarbeiter in das Projekt einbezogen.

### **5.2 Zusammenarbeit mit anderen Institutionen**

Die am Netzwerk SED-Archivgut sowie an der Ausweitung um die FDGB-Bestände beteiligten Archive werden in die Planungen und Entwicklungen einbezogen.

### **5.3 Apparative Ausstattung**

Die apparative Ausstattung wird gestellt.

### **5.4 Laufende Mittel für Sachausgaben**

Laufende Sachaufwendungen werden übernommen.

### **5.5 Sonstige Eigenleistungen und Voraussetzungen**

Die Projektmitarbeiter erhalten Arbeitsplätze mit üblicher Ausstattung sowie Internet- und e-mail-Anschlüssen. Arbeitsplatzkosten incl. Kommunikationskosten und Porto werden übernommen.

## **6 Erklärungen**

Ein Antrag auf Finanzierung dieses Vorhabens wurde bei keiner anderen Stelle eingereicht. Wenn ich einen solchen Antrag stelle, werde ich die Deutsche Forschungsgemeinschaft unverzüglich benachrichtigen.

## **7 Unterschrift**

Koblenz, 31. Juli 2006 / Berlin, 22. August 2006