

Förderantrag

Gegenstand:

Entwicklung von standardisierten Werkzeugen mit EAD, EAC und METS für unterschiedliche Archivgutarten und verschiedene Erschließungsebenen
auf der Basis des im <daofind>-Projekt Nov 2004 bis Juni 2006 entwickelten MEX-Werkzeugsets.

Zusammenfassung:	1
Ziele.....	2
- Beständeübersichten.....	3
- Findmittel für Sachakten, Standardversion.....	4
- Findmittel für Sachakten, erweiterte Version	4
- Findmittel für Nachlass-Bestände.....	5
- Findmittel für Personalakten und personenbezogene Unterlagen.....	5
- Anzeigesystem für digitalisierte Akten.....	6
Die Software-Grundlage der Tool-Entwicklung.....	7
Das METS-Anzeigesystem.....	13
Ergebnisse.....	17

Zusammenfassung:

Mit Unterstützung der Andrew W. Mellon Stiftung konnte im Rahmen des Projektes <daofind> der MEX Werkzeugkasten als Prototyp eines neuartigen archivischen Werkzeugs zusammen mit einer Pilotanwendung eines Präsentationsmodells für schriftliches Archivgut entwickelt werden, womit unter Nutzung der internationalen Standards EAD, EAC und METS die Darstellung von digitalisiertem Archivgut im Rahmen von Online-Findmitteln auf den Ebenen der Verzeichnungseinheit und des Bestands ermöglicht wird. Beide Ergebnisse haben die Machbarkeit dieses Ansatzes gezeigt. MEX soll nun als Entwicklungsplattform für spezialisiertere Werkzeuge genutzt werden.

Die Zielsetzung des neuen Projekt besteht darin, auf der Basis des MEX-Prototypen für den Werkzeugkasten spezialisierte Werkzeuge für konkrete Aufgaben zu entwickeln, um bei geringem Ressourcenverbrauch und mit reduziertem Einsatz menschlicher Arbeitskraft für Routineaufgaben die Bereitstellung einer kritischen Masse von digitalisiertem Archivgut im Internet zu fördern und auszuweiten. Die mit diesen Werkzeugen hergestellten und bearbeiteten Findmittel, die mit digitalen Reproduktionen angereichert sind, sind vollständige Webpräsentationen. Sie können direkt für die Bereitstellung auf einer Webseite wie für die Integration in Suchmaschinen wie Midosasearch, die im Bundesarchiv für die übergreifende Recherche und eine strukturierte Navigation eingesetzt wird, ebenso wie für die Datenlieferung an andere zentrale Zugangsplattformen genutzt werden. Die Werkzeuge werden mit eingebauten Style Sheets für standardisierte Präsentationsmodelle und einer HTML-Vorschau-Funktionalität bereit gestellt. Da die Anwender gültige XML-Dateien herstellen und sie bearbeiten, sind die Ergebnisse offen für andere Präsentationsformaten und andere Style sheets, die für die Präsentation gewählt werden können.

Ziele

Der Prototyp des MEX Werkzeugkastens basiert auf den drei Standards EAD, EAC und METS und erzeugt gültige XML-Dokumente. Er realisiert die vollständige Flexibilität des jeweiligen XML-Standards. Der MEX-Editor besteht aus leistungsfähigen und hochkomplexen Werkzeugen. Die große Flexibilität und Vielfalt möglicher Anwendungen geht jedoch einher mit der Anforderung, ständig Auswahlen zu treffen. Das neue Projekt zielt auf die Reduzierung der Komplexität durch die Entwicklung eines Sets von speziellen Werkzeugen für typische Arbeitsszenarien. Ein Teil der Komplexität wird bereits durch wichtige Entscheidungen im Vorfeld der Erschließung aufgehoben.

Wenn EAD für unterschiedliche Typen von Findmitteln mit Bezug auf unterschiedliche Materialien oder verschiedene Aggregierungsstufen genutzt werden kann, werden diese normalerweise nicht nach Arbeitsbeginn in einem EAD-Dokument erneut geändert. Aus diesem Grund kann die Auswahl ohne Flexibilitätsverlust auch bereits vor dem Beginn der Erschließung vorgenommen werden. Genauso kann die Entscheidung, ob eine Person, eine Körperschaft oder eine Familie als Urheber der Unterlagen mit einem

EAC-Dokument beschrieben werden soll, vor dem Beginn der Erfassung der Daten getroffen werden. Die Unterschiede zwischen den Standard-Werkzeuge resultieren aus unterschiedlichen Subsets der entsprechenden Standards, aus den speziellen Beschriftungen auf dem Bildschirm, bei denen zwischen verschiedenen Sprachen umgeschaltet werden kann, und aus angepassten Transformations-Style-Sheets für die HTML-Präsentationen.

Die standardisierten Werkzeuge werden die Einfügung von Links zwischen XML-Dateien unterstützen, die bei der kombinierten Präsentation benötigt werden, z. B. von EAD und EAC im Fall von Beständeübersichten oder von EAD und METS für in Findbücher integrierte Anzeigesysteme für Digitalisate.

Die folgenden sechs Spezialwerkzeuge werden im Rahmen des Projekts hergestellt:

- für Beständeübersichten auf der Bestandsebene,
- für Findbücher zu Sachakten auf der Ebene der Verzeichnungseinheiten in einer standardisierten und
- für Findbücher zu Sachakten auf der Ebene der Verzeichnungseinheiten in einer erweiterten Fassung,
- für Findbücher zu Nachlässen,
- für Findbücher zu Personalakten und personenbezogenen Unterlagen und
- für ein integriertes Anzeigesystem für digitale Reproduktionen in Online-Findbüchern.

Standard Style Sheets für Präsentationen, ähnlich denen, die bereits für MidosaXML Online-Findbücher im Einsatz sind, werden in die Werkzeuge integriert, um Findbücher verschiedener Typen erstellen zu können. Das integrierte Anzeigesystem für Digitalisate wird aus einem Erfassungswerkzeug bestehen, mit dem alle Metadaten eingegeben werden können, die für die Definition der inneren Struktur eines jeden digitalen archivischen Objekts, das jeweils einer Verzeichnungseinheit entspricht, benötigt werden, ergänzt durch ein Style Sheet zur Erstellung eines spezialisierten Präsentationsmodells, ähnlich dem, das in der Pilotanwendung des daofind-Projekts eingesetzt wird.

- Beständeübersichten

Beständeübersichten enthalten zusammenfassende Übersichten über alle Bestände einer archivischen Einrichtung einschließlich ihrer internen Verknüpfungen in einer Art

Klassifikationsschema, der sogenannten Tektonik. Sie können mit einer Kombination von EAD und EAC aufgebaut werden. Dabei wird EAD für die Beschreibung der Struktur mit Informationen über die Bestände selbst verwendet. EAC in den drei Varianten für Körperschaften, Familien und Personen wird für Informationen über die Herkunftsstellen mit für jede Variante angepassten Untergruppen von Elementen eingesetzt. EAC bietet Datenelemente, die für die Ansetzungskontrolle für offizielle Bezeichnungen und andere verwendete Namen und Bezeichnungen genutzt werden können und ermöglicht die Beschreibung von Kompetenzen und Zuständigkeiten sowie ihrer Entwicklung im Laufe der Zeit. In größeren Einrichtungen können die EAC-Daten dezentral von archivischen Fachkräften gepflegt werden, während das EAD-Dokument mit der Tektonik als Rückgrat zentral gepflegt wird. Das Werkzeug wird Subsets von EAD und EAC einsetzen und angepasste Beschriftungen sowie ein Style Sheet nutzen, das die für die Präsentation des jeweiligen Dokuments notwendige Kombination von EAD- und EAC-Daten erzeugt.

- Findbücher für Sachakten, Standardversion

Die Beschreibungselemente für Standard-Findmittel sind reduziert auf die wesentlichen Daten über eine Verzeichnungseinheit, während allgemeine Informationen über einen gesamten Bestand in einer Einleitung oder in Informationen über die Verarbeitung auf der obersten Strukturebene, gegeben werden können. Dieses Werkzeug wird nur die Grundelemente für Findmittel anbieten, so dass es auch für den Einsatz an Arbeitsplätzen von weniger qualifiziertem Personal geeignet ist. Die entstehenden XML-Dokumente können mit der erweiterten Version geöffnet werden, wenn sich die Notwendigkeit ergeben sollte, mehr und differenziertere Information an beliebiger Stelle innerhalb des Dokuments einzufügen. Diese Version nutzt einen reduzierten Satz von EAD-Elementen und die normalen Beschriftungen. Es wird kein besonderes Style Sheet benötigt.

- Findbücher für Sachakten, erweiterte Version

Die erweiterte Version bietet mehr Möglichkeiten für die Erschließung einzelner Verzeichnungseinheiten, für zusätzliche Informationen und für die Dokumentation der Bearbeitung und der Bewertungsentscheidungen auf unterschiedlichen Hierarchieebenen in der Struktur eines Findmittels. Im Gegensatz zur Standardversion

werden mehr Wahlmöglichkeiten angeboten, so dass ausgebildetes und erfahrenes Personal anspruchsvollere Erschließungsmethoden anwenden kann. Diese Version setzt ein größeres Set von EAD-Elementen mit der Standardbeschriftung und dem Standard-Style Sheet wie bei MidosaXML-Online-Findbüchern ein, das z. B. Möglichkeiten zur Erfassung der Bearbeitungsdokumentation oder Informationen zu Bewertungsentscheidungen anbietet.

- Findbücher für Nachlass-Bestände

Die Erschließung von Nachlass-Beständen erfordert oft Raum für biographische Informationen über die Person des Nachlassers und über die Aktivitäten während der Zeit, in der die Unterlagen des Nachlasses entstanden und zusammengeführt wurden. Das unterscheidet diese Findbücher von solchen für Sachakten. Daher wird ein spezielles Subset von EAD-Tags benötigt. Noch wichtiger ist die Beschriftung der Felder für die verschiedenen Elemente des Findbuchs in diesem Werkzeug, damit bei der Erschließungsarbeit die angemessenen Platzierungen für das während der Arbeit am Bestand sich erweiternde Wissen über die Person ebenso wie über die Entstehung und die Entwicklung des Nachlasses bereit stehen. Diese Version wird voraussichtlich ein angepasstes EAD-Subset für die Beschreibung des Bestands und ein EAC-Subset für die Nachlasser selbst verwenden. Die Kombination von EAD und EAC für in dem Bestand erwähnte Personen wird in der Form integriert sein, dass die EAC-Daten gesondert herausgezogen, mit Normdateien abgeglichen und zur Vervollständigung existierender Normdateien verwertet werden können. Eine eigene Beschriftung und ein angepasstes Style Sheet für die Kombination von EAD und EAC ist erforderlich.

- Findbücher für Personalakten und personenbezogene Unterlagen

Personalakten entstehen durch die Aktivitäten im Rahmen der Verwaltungserfordernisse mit Bezug auf einzelne Personen oder Personengruppen. Diese Daten sind gewöhnlich nach Namen oder anderen Identifikationsmerkmalen, wie etwa dem Geburtsdatum, in der Form von Serien geordnet. Solche Unterlagen können in einer stark standardisierten Weise erschlossen werden. Zusätzlich zum Titel der Akten werden weitere Merkmale für die Erschließung benötigt und oft sind weitere Angaben aus den Unterlagen für das Findmittel gewünscht. Hier kann ebenfalls ein reduziertes EAD-Subset eingesetzt werden. EAC wird mit den EAD-Daten verbunden

werden und die Möglichkeit zur Nutzung einer Schnittstelle zu Normansetzungen eröffnen. Dies Tool wird wiederum ebenso ein gesondertes Set von EAD-Elementen benötigen wie ein angepasstes Set von EAC-Elementen mit einer eigenen Beschriftung und einem besonderen Style Sheet mit mehr Informationskategorien, wie z. B. biographischen Angaben zu Personen.

- Integriertes Anzeigesystem für digitalisierte Akten

Digitale Images werden in Online-Findbücher eingebaut und sind mit Hilfe des auf der METS-Datei mit allen erforderlichen Metadaten basierenden Anzeigesystems anzusehen. Die METS-Datei verbindet alle Images einer Verzeichnungseinheit zu einem digitalen archivischen Objekt mit der Möglichkeit, dessen innere Struktur darzustellen. In jede METS-Datei können automatisch die Adressen der einzelnen Bilder eingelesen werden. Das zugehörige Style Sheet erzeugt eine HTML-Präsentation, die mit der Verzeichnungseinheit in dem entsprechenden Findbuch so verlinkt ist, so dass sie mit Anklicken geöffnet wird und die Bilder auf zwei Ebenen anbietet. Die strukturierenden Möglichkeiten von METS können auf Wunsch genutzt werden, um auf der oberen Ebene die strukturelevanten Seiten anzuzeigen und damit eine Orientierung für eine Navigation durch die digitalisierten Unterlagen zu geben.

Das Präsentationsmodell für Online-Findbücher, das zusammen mit dem MidosaXML entwickelt wurde, wird die Standardpräsentation sein, die mit dem Werkzeug ausgeliefert wird. Die Verzeichnungseinheiten werden Verknüpfungen zu METS-Dateien mit den Digitalisaten anbieten. Das Online-Findbuch wird in einem Set von drei Frames angezeigt. Ein linker Frame zeigt eine auffächerbare Strukturübersicht wie ein Inhaltsverzeichnis in einem Buch an, mit Links zum Text des Dokuments, das im rechten Hauptframe durch Klicken auf die entsprechende Überschrift angezeigt wird. Die Gliederungsgruppe, die auf diese Weise aufgerufen wird, wird vollständig wie ein Kapitel in einem e-Book in einem verschiebbaren Hauptrahmen angezeigt. Oberhalb des Hauptrahmens wird die vollständige Gliederung von der gewählten Gruppe an aufwärts angezeigt. Dieser Frame hat eine ähnliche Funktion wie Kopfzeilen in gedruckten Büchern und unterstützt die Orientierung. Alle drei Frames sind synchronisiert. Das heißt, dass Seitenköpfe und der Navigationsbaum bei einem Wechsel des Inhalts des Hauptfensters aktualisiert werden und umgekehrt. Dieses Präsentationsmodell ermöglicht die Nutzung und Kombination von vier verschiedenen

Recherchestrategien in einer Suche. Das sind eine Navigation in der Struktur, eine Volltext-Suche, eine blätternde Suche durch das gesamte Findmittel und die Nutzung von Indexbegriffen, was unmittelbar zu dem in der Referenz bezeichneten Ort führt.

Dieses Standardmodell für Findbücher wurde als Grundlage für die Integration von Bildern verwendet. Wenn eine METS-Datei mit Metadaten über Images verfügbar ist und ihre Adresse im <dao>- Element der EAD-Datei angegeben ist, wird während der Ausgenerierung der HTML-Datei ein Link mit einer in der EAD-Datei angegebenen Beschriftung erzeugt, der in der Präsentation ein neues Fenster mit der Anzeige für die Images öffnet.

Die Anzeige mit dem Pilot-Präsentationsmodell wird auf der Website des daofind-Projekts (www.daofind.de) mit drei exemplarischen Findmitteln vorgeführt, die mit 30.000 Images aus drei Beständen angereichert wurden. Zwei vollständige Bestände und ein Teil eines dritten Bestandes wurden als Versuchsumgebung für die Pilotanwendung ausgewählt. Die ersten beiden Bestände stammen aus zwei Sekretariaten des Zentralkomitees der sozialistischen Partei (SED) der früheren DDR mit Unterlagen aus den Jahren 1945/46 bis in die Mitte der 1950er Jahre, einer historisch sehr interessanten Periode während der Gründung des ostdeutschen Staates. Der dritte Bestand stammt aus der Zentralen Partei-Kontrollkommission der SED, einem Disziplinargericht innerhalb der Partei, mit mehr als zehn Akten über allgemeine politische Angelegenheiten. Alle Digitalisate wurden vom Mikrofilm hergestellt, der zuvor für Zwecke der Bestandserhaltung angefertigt worden war. Die Findmittel sind in EAD kodiert. Jede Verzeichnungseinheit ist mit der entsprechenden METS-Datei über den Button „Akte einsehen“ verbunden.

Die Software-Grundlage der Werkzeug-Entwicklung

Die Entwicklung der Werkzeuge wird vom MEX-Prototypen des daofind-Projektes ausgehen. Die Software ist konzipiert als Plug-In für die Open-Source-Java-Entwicklungsumgebung Eclipse. Eclipse basiert selbst auf Java und ist daher plattformunabhängig. Es läuft auf verschiedenen Betriebssystemen wie Windows, MAC-OS und Linux. Eine vollständige Anwendungsfähigkeit unter dem Apple Betriebssystem hängt von der weiteren Entwicklung der Eclipse-Plattform ab, die für die nähere Zukunft erwartet werden darf.

Eclipse stellt eine sehr innovative Umgebung zur Verfügung, die von IBM entwickelt wurde, das den Quellcode für die Open-Source-Gemeinschaft offengelegt hat. Es entwickelt sich sehr rasch und besitzt ein hohes Potenzial für die weitere Entwicklung der Internet-Technologien. Es wird in ständig steigendem Maße genutzt und ist daher ausreichend stabil, um als Grundlage eines professionellen Arbeitswerkzeugs zu dienen.

Der Prototyp für die spezifischen Werkzeuge umfasst bestimmte Funktionen, die normalerweise nicht für XML-Editoren zur Verfügung stehen wie eine integrierte HTML-Ansicht und eine Funktion für die schnelle Erzeugung eines Sets von HTML-Dateien, die für die Internet-Präsentation eines XML-Dokuments benötigt werden. Diese Funktionalitäten werden an die Werkzeuge weitergeben, die im Rahmen des neuen Projektes entwickelt werden.

Der Prototyp MEX wird als Plug-In zu Eclipse installiert. Sobald begonnen wird, eine XML-Datei zu erzeugen, die einem der Standards entspricht, wird das zugehörige XML-Schema in den Arbeitsbereich kopiert. Der Editor öffnet ein XML-Dokument mit gerade dem für ein entsprechend dem Standard gültiges XML-Dokument notwendigen Satz an Elementen. Die übrigen Elemente können bei Bedarf hinzugefügt werden. Alle drei Editoren des MEX-Werkzeugs werden nach dem Grundprinzip entwickelt, dass ihre Anwendung keine Kenntnisse von XML erfordern soll, und sie stattdessen auf allgemeinen Fachkonzepten und entsprechender Terminologie beruhen sollen und in unterschiedliche technische Umgebungen integriert werden können.

Das Ziel ist, dass die Werkzeuge für die alltägliche Arbeit am archivischen Schreibtisch genutzt werden können und es erlauben, die Arbeit in einer neuen, stärker ergebnisorientierten Weise zu erledigen. Für die Präsentation archivischer Arbeitsergebnisse im Internet soll kein speziell ausgebildetes Personal notwendig sein. Vielmehr sollen integrierte Arbeitsprozesse es ermöglichen, mit denselben, einmal erfassten Daten ebenso Internet-Präsentationen zu erzeugen wie Offline-Publikationen, also CDs, DVDs oder gedruckte Findmittel, ohne dass manuelle Transformationen oder Überarbeitungen für andere Publikationsformen notwendig sind.

MEX arbeitet auf der Dokumentenebene, das heißt auf einer dokumentenbezogenen Ebene oberhalb der technischen XML-Ebene. Wer das Tool einsetzt, erzeugt Dokumente, etwa Findbücher, entsprechend den eigenen fachlichen Kompetenzen und

unter Verwendung einer Fachsprache, die präzise und eindeutig ist. Unter der sichtbaren Oberfläche werden die Konzepte und Begriffe der archivischen Fachsprache in das entsprechende XML-Schema umgesetzt. Eine Eins-zu-eins-Beziehung zwischen den Begriffen der natürlichen Sprache und den XML-Elementen wird dazu nicht benötigt. Die Konzepte und Begriffe können in einzelne EAD-konforme Elementbezeichnungen, also Tags, in Tag-Kombinationen oder in Kombinationen von Tags und Attributen, wenn nötig mit speziellen Werten, übersetzt werden. Diese Flexibilität eröffnet die Möglichkeit, die auf dem Arbeitsbildschirm benutzten Beschriftungen leicht zu ändern, wenn sie die gleiche Funktion, aber eine abweichende Bezeichnung haben, wie etwa bei der Erschließung von bestimmte Archivguttypen, z. B. Nachlässen im Unterschied zu Unterlagen staatlicher Behörden. Auf die gleiche Weise kann die Sprache geändert werden, um Versionen für unterschiedliche Nationalitäten einzustellen. Die speziellen Werkzeuge, die im Rahmen des neuen Projekts entwickelt werden, werden verschiedene Fachsprachen nutzen, die für verschiedene Anwendungen benötigt werden.

Die Informationen über den Gebrauch der Elementbezeichnungen und ihre Beschriftungen auf dem Bildschirm sind in einem eigenen XML-Dokument erfasst und dokumentiert, das alle Definitionen für die jeweilige Anwendung jedes Elements oder Attributs des entsprechenden Schemas enthält. Diese Definitionsdateien für die Übersetzung der Fachsprache in die Sprache des XML-Standards bieten gleichzeitig die Möglichkeit, eingeschränkte Subsets der Standards zu definieren, die für besondere Formen von Findmitteln angeboten werden. Bestimmte Elemente und Attribute und ihre Werte können außen vor gelassen werden, wenn sie nicht benötigt werden. Dann werden sie auf dem Arbeitsbildschirm nicht angeboten und werden damit auch keine Auswahlentscheidungen erfordern, die für den speziellen Findmitteltyp nicht notwendig sind.

Die XML-Dateien mit den Definitionen werden zum Aufbau eines Arbeitsbildschirms verwendet. Daher stellt der Arbeitsbildschirm nur genau das Subset des Standards dar, das für die Arbeit in der konkreten Arbeitssituation benötigt wird. Trotzdem sind die erschließenden archivischen Fachkräfte immer in der Lage, genau die Elemente eines Findbuchs zu nutzen, die für den speziellen Erschließungsfall erforderlich sind. Jedes mit dem Werkzeug erstellte Dokument wird gegen das jeweilige XML-Schema auf Gültigkeit validiert, wenn es gespeichert wird.

Eclipse bietet die Möglichkeit unterschiedlicher Bildschirmdarstellungen. Im Mex-Editor zeigt der Bildschirm mit der Bearbeitungsoberfläche im mittleren Frame eine Ansicht, in der Texte eingegeben und Veränderungen vorgenommen werden können. Sie umfasst die mit Inhalt gefüllten Felder in strukturierter Darstellung. Im Werkzeugkasten auf der rechten Seite werden Abschnitte und Elemente der Struktur der zugehörigen Findbücher angeboten. Mit Klick auf eine Bezeichnung öffnet sich das entsprechende Feld im mittleren Frame, dem Arbeitsbereich. Hier kann dann der Inhalt eingegeben und bearbeitet werden. Auf der linken Seite kann in einer Strukturansicht verfolgt werden, wie die Struktur wächst und sich verändert. Andere Ansichten zeigen die Ergebnisse der Validierung oder Hilfetexte zu den ausgewählten Elementen. Alle Ansichten können auf Wunsch geöffnet und geschlossen werden. In der Dokumentenansicht im mittleren Frame kann mit Hilfe von Reitern zwischen der Editorenansicht, einer HTML- und einer XML-Ansicht umgeschaltet werden, so dass die einzelnen Handlungen und ihre Ergebnisse umgehend verfolgt werden können.

Die Werkzeuge besitzen eine Importfunktionalität, die jedes Dokument lesen kann, das zu den jeweiligen Standards konform ist und eine gültige Instanz darstellt. Die Elemente und Attribute, die in der gerade eingestellten Definitionsdatei nicht angegeben sind, werden mit ihrer in Klammern gesetzten, aus dem Schema ausgelesenen Bezeichnung angezeigt, womit angedeutet wird, dass sie in der aktuellen Umgebung nicht bearbeitet werden können. Trotzdem kann diese Datei mit dem integrierten Style Sheet nach HTML transformiert werden. Neue Elemente, wie Links zu Bildern oder anderen XML-Dateien, können ebenfalls hinzugefügt und die Datei kann danach als gültige XML-Datei mit den neu hinzugefügten Feldern abgespeichert werden. Die Style Sheets sind in XSLT 2.0 erstellt und erlauben die Erzeugung einer HTML-Präsentation sowie die Erstellung einer Text-Datei als Druckmanuskript.

Die Werkzeuge, die im Rahmen des neuen Projekts erstellt werden, werden für den Einsatz in typischen Situationen in der archivischen Erschließungsarbeit vorbereitet sein. Sie werden den Editor enthalten, der wie ein XML-Editor arbeitet, aber fachsprachlich bezeichnete Elemente des jeweils bearbeiteten Erschließungsdokuments, also eines Findbuchs oder einer Beständeübersicht, benutzt. Sie werden über eine Import-Funktionalität verfügen, die für jedes dem entsprechenden Standard konforme Dokument offen sein wird. Und sie werden Standard-Style Sheets für Online- und Druckfindmittel enthalten. Das Layout der Online-Findmittel wird sich je nach dem

speziellen Typ leicht unterscheiden. Daher werden die speziellen Werkzeuge mit entsprechend voneinander abweichenden Style Sheets ausgeliefert. Diese Werkzeuge werden kostenfrei vertrieben und der Quellcode wird offen gelegt werden.

Die Flexibilität zur beliebigen Definition weiterer Subsets mit zugehörigen Beschriftungen der Arbeitsoberfläche wurde im Rahmen des ersten Projekts entwickelt und durch einen zusätzlichen Einsatz von Mitteln des Bundesarchivs ermöglicht. Die Softwareentwicklungsfirma erarbeitete sich innovative Programmierungstechniken für diesen Teil und hat selbst eigene Ressourcen investiert, um innovative Kompetenzen für die eigene Geschäftstätigkeit zu erwerben. Aus diesem Grund geht der Antragsteller davon aus, dass die Funktion für die freie Definition der benötigten Subsets lizenziert bleibt, während alle sonstigen Programmdateien und Style Sheets weitergegeben und offen gelegt wird. Die frei abgebbaren Werkzeuge schließen eine unbeschränkte Import-Funktionalität ebenso mit ein wie Standard-Style Sheets und die Offenheit für die Nutzung anderer Style Sheets.

Die neuen Werkzeuge werden unabhängig und autonom einsetzbare Instrumente für die Erstellung von XML-Dateien sein, die den Standards EAD, EAC und METS entsprechen. Die XML-Dateien werden mit jedem anderen Editor bearbeitet werden können. Die notwendigen Elemente werden angeboten und können nach Bedarf ausgewählt werden. Die Auswahl und die Bildschirmdarstellung der Tags für die unterschiedlichen Anwendungen werden festgelegt, indem die speziellen Definitions-Werkzeuge des Prototyps verwendet werden, die lizenziert bleiben werden. Diese Werkzeuge sind für den Einsatz der einzelnen Werkzeuge bei der archivischen Arbeit nicht erforderlich. Die Editoren werden so ausgelegt sein, dass die in XML-kodierten Definitionsdateien auch ohne das Definitionswerkzeug manuell verändert werden können.

Die MEX-Editoren können auf jeder Plattform (Windows, Linux oder Apple Macintosh) und auf jedem Computer eingesetzt werden, da sie in Java geschrieben werden und auf der Java-Entwicklungsplattform Eclipse basieren. Sie können als Stand-alone-Systeme eingesetzt werden oder in Netzwerke integriert werden und nutzen den üblicherweise bereitgestellten Speicherplatz. Sie benötigen im Hintergrund keine Datenbank-Management-Systeme, sondern nutzen stattdessen Dateisysteme für die Datenhaltung. Die Ergebnisse in Form von automatisch erzeugten HTML-Dateien können ohne weitere Adaptierung in Websites integriert werden. Daher erfordern sie wenig bis keine

IT-Personalkapazitäten für ihre Anwendung, für die Präsentation der Findmittel und Bilder im Internet und für Management und Pflege der Daten.

Die EAD-Bearbeitungs-Funktionalität von MEX erlaubt die Nutzung des <c>-Elements für die Erzeugung und Gliederung hierarchischer Strukturen entsprechend den Definitionen des EAD-Schemas. Die Struktur kann durch die Verschiebung von Gliederungsüberschriften- oder Verzeichnungseinheiten verändert werden. Der EAC-Editor kann für die Erfassung aller Informationen über Herkunftsstellen einschließlich unterschiedlicher Bezeichnungen und der Zeiträume, in denen diese verwendet wurden, eingesetzt werden. Der Herkunftsstellen können Personen, Familien oder Körperschaften sein. Kompetenzen können aufgelistet und beschrieben werden zusammen mit dem Zeitraum, in dem sie galten. EAC ist geeignet für die Teile von Beständeübersichten, in denen die Bestandsbeschreibungen erfasst werden und die Provenienz der Bestände beschrieben wird.

Für das Werkzeug zur Erzeugung von Beständeübersichten wird je eine EAC-Datei zur Beschreibung jeder einzelnen Herkunftsstelle genutzt. Sie wird mit der EAD-Datei verknüpft sein, die für die Anzeige der Struktur einer Beständeübersicht benutzt wird und die Tektonik repräsentiert. Die Integration von EAD und EAC soll während der Erstellung der HTML-Ansicht im Browser an Hand eingefügter Verknüpfungen realisiert werden. Damit könnte die Pflege der Inhalte dezentral erfolgen. Diese Funktion soll im Rahmen des Projekts entwickelt werden. Inwieweit sich die Anwender der Nutzung verschiedener Standards dabei bewusst sein sollten, ist noch nicht entschieden. Einfachheit der Nutzung als auch die Angabe notwendiger Hinweise zur Orientierung sollten bei der Konzepterstellung für dieses Werkzeug Leitlinie sein.

Die Bearbeitungsfunktion für METS-Dateien kann verwendet werden, um halb-automatisch gültige METS-Dateien zu erzeugen, die unmittelbar mit EAD-Findmitteln verknüpft werden können. Alle Dateiadressen von Digitalisaten, die innerhalb eines METS-Dokuments zusammengefasst werden sollen, können automatisch in einen Dateigruppen-Abschnitt des METS-Dokuments durch Auswahl des entsprechenden Ordners eingetragen werden.

Gleichzeitig werden eine Strukturübersicht erstellt. Danach kann die Datei bereits mit einem EAD-Findmittel verknüpft werden. In ihr sind alle digitalen Dateien aus einer Verzeichnungseinheit zu einem digitalen Archivobjekt zusammengefasst und können

über eine Verknüpfung aus dem entsprechenden Online-Findmittel aufgerufen werden. Auf Wunsch können weitere Metadaten manuell hinzugefügt werden mit denen die Präsentation der Bilder gesteuert wird.

Das zugrunde liegende Konzept geht davon aus, dass ein METS-Dokument einer Verzeichnungseinheit in einer EAD-Datei entspricht und dass es alle dazugehörenden digitalen Dateien zusammenfasst, falls erforderlich, in mehreren Formatgruppen, z. B. für unterschiedliche Auflösungen oder für die Integration von OCR-Transkriptionen zu den Bildern.

Das METS-Anzeigesystem

Da keine speziellen Anzeigesysteme für Schriftgut bei anderen Präsentationen im Internet zu finden waren, hat das daofind Projekt ein eigenes Präsentationsmodell konzipiert, das besonders an schriftliches Archivgut angepasst ist. Dieses spezielle Präsentationsmodell für METS-Dateien mit allen Metadaten der digitalen archivischen Objekte, die jeweils einer Verzeichnungseinheit auf der unteren Ebene der EAD-Hierarchie entsprechen, erlaubt einerseits die schnelle Einschätzung, ob eine Seite detailliert betrachtet werden muss und gibt andererseits eine bessere Orientierung während des Durchsehens von umfangreichen Akten, selbst wenn sie etwa 500 Seiten oder mehr enthalten. Die digitalen archivischen Objekte, die eine METS-Datei abbildet, bestehen aus einer bestimmten Anzahl von Dokumenten, die zu einander in Beziehung stehen wie eine Kette von Kommunikationsereignissen, mit eingehenden Schreiben und Konzepten für Antworten oder Berichte und Reaktionen, Entwürfe und den gleichen Fall betreffende Ausfertigungen. METS wird oft im Internet verwendet, um Bilder oder einzelne Seiten, wie etwa handschriftliche Briefe, Postkarten oder ähnliche Stücke zu präsentieren. Gewöhnlich werden die Bilddateien zunächst als stark verkleinerte Vorschaubilder angezeigt, die durch Mausklick in einer vergrößerten Darstellung geöffnet werden können. Diese Darstellungsform ist nicht optimal für eine Zusammenstellung von 100 bis 200 oder gar mehr Bilddateien von einem Aktenstück mit kaum vorhandenen äußeren Unterschieden. Die Absicht hinter dem Konzept eines neuen Präsentationsmodell war es, eine reduzierte Reproduktion zu zeigen, die dennoch ausreichend Informationen enthält, um abzuschätzen, ob es sich lohnt, darauf zu klicken, um die Vergrößerungsform zu erhalten. Die äußere Form der Dokumente, die in einer kleinen Bilddatei erkannt werden kann, gibt nicht genügend Informationen,

um diese Entscheidung zu treffen. Aber das obere Drittel eines Dokuments enthält oft Informationen über den Absender, über die Datierung sowie den Betreff und zeigt den Anfang des Textes. Dieses Seitendrittel wird deshalb angezeigt. Es erscheint hilfreicher als etwa stark verkleinerte Bilder, um die Relevanz der ganzen Seite und von folgenden Seiten für das eigene Informationsbedürfnis abzuschätzen.

Die zweite Überlegung bestand darin, dass es für viele umfangreiche Dokumente in einer Verzeichnungseinheit, wie z.B. Protokolle oder Berichte, ausreicht, die erste Seite anzusehen, um zu entscheiden, ob das ganze Dokument wichtig ist. Daher erlaubt es eine Auswahl dieser ersten Seiten und anderer strukturell relevanter Seiten der Akte, etwa solche, die eine neue Handlung oder eine neue Kommunikationskette initiieren, den Lesern, den Nutzen einzuschätzen, ob sie den ganzen Bericht oder das vollständige Protokolls einsehen wollen.

Deshalb enthält das METS-Anzeigesystem zwei Ebenen. Die erste ist als Orientierungsebene gedacht, die das obere Drittel derjenigen Seiten enthält, die als nützlich ausgewählt wurden, um den Lesern eine erste Einschätzung der Relevanz dieser Unterlagen für die eigene Forschungsfrage zu bieten. Dieses obere Drittel wird in verschiebbaren Fenstern auf einer Seite angezeigt und ein Mausklick öffnet die Vollansicht der Bilddatei in der Blätteransicht, wo ein zweiter Klick die nächste Seite öffnet. Oben angegebene Seitenzahlen sowie Pfeilsymbole unterstützen schnelles Betrachten. Ein Link zurück zur Orientierungsebene ermöglicht die Rückkehr zum Ausgangspunkt. Diese Darstellungsform wird direkt aus METS erzeugt ohne eine zwischengeschaltete Umwandlung der Daten.

Die Auswahl der Seiten für die Orientierungsebene erfolgt durch die erschließende Fachkraft nach der Erstellung der vorhandenen METS-Datei mit der automatisierten Integration aller Adressen aller Dateien und der Erstellung einer Strukturansicht. Sie geschieht durch die manuelle Eingabe bestimmter Attribut-Werte für das zweite <div> Element in der Strukturansicht der METS-Datei. In der Arbeitsansicht wird die Beschriftung „Schriftstückform“ zur Auswahl von einem aus acht Typen, z.B. „eingehendes Schreiben“, angeboten. Außerdem kann eine Bezeichnung eingefügt werden, die als Beschreibung des Bildes erscheint. Im XML-Dokument wird die folgende Kombination von Elementen und Attributen für diesen Begriff in der Arbeitsansicht hergestellt, indem sie der integrierten Definition auf der Basis des METS-Schemas folgt: <structMap><div><div>TYPE=„eingang“. Das erste, hierarchisch obere

<div>-Element in der Strukturansicht kann für die Auswahl anderer Präsentationsmodelle für andere Typen von Verzeichnungseinheiten ausgewählt werden.

Das Konzept für dieses Arbeitswerkzeug sowie für das Präsentationsmodell ist völlig neu. Bei Vorführungen erhielt es bisher viel Zustimmung. Das Präsentationsmodell kann auf der Webseite des daofind-Projektes mit Findbüchern getestet werden, in die 30.000 Bilddateien eingebunden sind. Die Präsentationsseiten werden mit weiteren Kontextinformationen aus der korrespondierenden Verzeichnungseinheit versehen werden. Auf der Webseite ist auch eine englischsprachige Folienpräsentation verfügbar, die die Arbeitsansicht zeigt. Mit der Nutzung von METS können die darin erfassten Metadaten direkt für die HTML-Präsentation verwendet werden, die wiederum während des Bearbeitens der METS-Datei kontrolliert werden kann. Ein Profil für diese spezifische Verwendung von METS wird definiert und beim METS-Herausbergremium eingereicht werden.

Alle Werkzeuge, die mit dem neuen Projekt entwickelt werden, bieten besondere Funktionalitäten, die die Funktionen von einfachen XML-Editoren bei weitem übertreffen. Ein zentraler Aspekt besteht darin, dass die Arbeit auf der Ebene des Dokumentes und nicht auf der Ebene des XML-Schemas erledigt wird. XML-Elementbezeichnungen brauchen nicht gelernt zu werden. Sie werden nicht in der Arbeitsansicht angezeigt. Die stattdessen verwendeten Beschriftungen sind Begriffe, die der Fachsprache entstammen und in die entsprechenden Kombinationen von XML-Elementen und entsprechenden Attributen übersetzt werden. Wenn das Werkzeug mit einem der Standards gestartet wird, werden alle Elemente, die für eine gültige XML-Datei notwendig sind, eingegeben. Elemente können in dem Ausmaß verschoben werden, wie es das Schema erlaubt.

Von der Arbeitsansicht des Editors aus kann eine HTML-Listendarstellung und eine XML-Ansicht geöffnet werden. Allerdings können Änderungen nur in der Bearbeitungsansicht vorgenommen werden. Damit soll die Erhaltung der Konsistenz gesichert werden. Mit Schaltern oberhalb der Arbeitsansicht kann die hierarchische Struktur der Dokumente geöffnet oder geschlossen werden.

Der METS-Editor bietet zusätzlich zum EAD- und EAC-Editor während der Bearbeitung eine synchronisierte Bildansicht, in der die entsprechende Digitalisatdatei angezeigt

wird. Sie kann als Vollbildschirmansicht vergrößert werden, um durch das ganze digitale Archivobjekt zu blättern und wieder verkleinert werden, um weitere Metadaten zu einem Bild, insbesondere über die Struktur und zur Steuerung der Präsentation auf der Orientierungsebene, hinzuzufügen. Beim Weiterblättern ist diese Bildansicht ständig mit dem Editor der XML-Datei synchronisiert.

Die Programmierung der Software wird an die neuen Entwicklungen Eclipse 3.2 und der neuen Standardversion von XSLT 2.0 angepasst. Die Links zwischen den XML-Dokumenten, die erforderlich sind, um das Netz der HTML-Seiten für die Präsentation aufzubauen, werden den XLink Namensraum benutzen.

Die Arbeitsdokumente werden direkt im XML-Format erfasst, geändert und gespeichert. Eine Überführung in ein relationales Datenbankformat entfällt damit. Somit sind die Werkzeuge sehr schnell. Im Gegensatz zu XML-Editoren ist in den MEX-Editor sowohl ein HTML-Export integriert als auch ein Export in ein Textverarbeitungsformat. Der HTML-Export gestattet die sofortige Präsentation im Internet, die ohne Änderungen über den Server hochgeladen werden kann. Alle Dokumente können mit den anderen Dokumenten verknüpft und für die Darstellung eines Netzwerks einschließlich der Beständeübersicht genutzt werden, das über die mit einander verknüpften Findbücher Zugang zu den Bilddateien des Bestandes ermöglicht.

Ein weiterer Schritt wird die Ausweitung des Präsentationsmodells für digitalisiertes Archivgut durch eine Volltextsuche in den Bilddateien sein. OCR-Transkriptionen oder abgeschrieben Text wird dann für die Suche verwendet, und die Ergebnistreiber werden die entsprechenden Bilddateien aufrufen. Da der Text nur für die Suche verwendet und nicht angezeigt wird, ist eine größere Anzahl an Fehlern hinnehmbar, besonders wenn man die Tatsache berücksichtigt, dass eine Volltextsuche immer durch die strukturierte Navigation und die Verwendung der Querverbindungen unterstützt wird.

Während des daofind-Projektes hat das Bundesarchiv eine neue Suchmaschine für Online-Findbücher, Midosasearch, eingeführt. Sie basiert auf Lucene und erlaubt die übergreifende Suche über alle Findmittel. Sie kombiniert die Volltextsuche in den Findmitteln mit der strukturierten Navigation auf der Ebene der Beständeübersicht und des einzelnen Findbuchs. Das einzelne Findbuch kann damit entweder über die Volltextsuche oder über die Navigation in der Tektonik gefunden werden. Mehrere Findbücher können für übergreifende Suchen ausgewählt werden, und die Ergebnisse

führen direkt an die entsprechende Stelle im Findbuch oder in der Beständeübersicht, an der ein Suchbegriff verwendet wird. Diese Suchmaschine integriert bald direkt Findmittel, die in EAD kodiert sind. Im Moment sind 200.000 Verzeichnungseinheiten durchsuchbar. Im Frühjahr wird diese Suchmaschine der Hauptzugang zu allen On-line-Findmitteln des Bundesarchivs sein mit ungefähr einer Million Verzeichnungseinheiten in 6000 Beständen. Diese Suchmaschine wird als Basisversion kostenfrei über die Archivschule, Institut für Archivwissenschaft, Marburg, der zentralen Ausbildungsstätte für Archivare in Deutschland, verteilt. Die Serverversion der Suchmaschine für große Mengen an Verzeichnungseinheiten ist von der Entwicklungsfirma lizenziert. Die Archivschule ist ebenfalls bereit, mit dem Bundesarchiv die Verteilung der neuen Werkzeuge zu gewährleisten und sie sowohl in den Ausbildungs- als auch den Fortbildungskursen einzusetzen.

Das Projekt The Archivist's Toolkit (archiviststoolkit.org) scheint zudem einige Funktionalitäten für die Behandlung von normierten Begriffen und Ansetzungskontrolle mit Schnittstellen zu Normdatenbanken zu bieten, die eine interessante Ergänzung der zu entwickelnden Werkzeuge darstellen könnten. Da es wenig Erfahrung mit dieser Form von Erschließung in Deutschland gibt, die Notwendigkeit der Auseinandersetzung mit diesen Ansätzen jedoch akzeptiert wird, soll die Möglichkeit näher untersucht werden, bestimmte Funktionalitäten des Archivist's Toolkit Projektes für die neuen Werkzeuge nach- und weiterzunutzen. Dies erscheint deshalb besonders interessant, weil das Archivist's Toolkit mit EAC kompatibel ist. Mögliche Schnittstellen sollen untersucht werden. Der Hauptunterschied zwischen beiden Ansätzen jedoch besteht darin, dass MEX direkt mit den inneren Verknüpfungen und Strukturierungen von XML ohne Umsetzung in relationale Datenbankstrukturen arbeitet.

Ergebnisse

Die Ergebnisse des Projektes werden eine Reihe von Erschließungswerkzeugen für besondere archivische Bearbeitungsszenarios sein. Es ist nicht nötig, dass die archivischen Fachkräfte für die Erschließungsarbeit die Elementbezeichnungen der Standards erlernen. Sie können im gewohnten konzeptionellen und fachsprachlichen Umfeld arbeiten. Die im Editor eingesetzte Sprache verwendet Fachbegriffe und anstatt mit XML-Elementen arbeiten die Anwender mit den Elementen und Bestandteilen der Dokumente. Die Werkzeuge integrieren die Herstellung von Internet-Präsentationen in

die tägliche archivische Arbeit und können daher dazu beitragen, größere Mengen an Online-Findmitteln mit größeren Mengen integrierter Bilddateien in kürzerer Zeit als heute möglich herzustellen. Besonderes IT-Personal ist nicht erforderlich für die separate Aufbereitung archivischer Erschließungen für Internetpräsentationen, weil diese Arbeiten in die herkömmlichen Arbeitsabläufe integriert sind. Die Werkzeuge werden die Kosten für Digitalisierung von Archivgut, die hauptsächlich aus den Personalressourcen für die Erschließung und Eingabe der Metadaten bestehen, deutlich senken können.

Die spezialisierten Werkzeuge werden gültige XML-Dateien herstellen, die mit den Anforderungen der drei Standards übereinstimmen. Sie sind offen für die Präsentation mit verschiedenen Style-Sheets für HTML Webdarstellungen der Bestände ebenso wie für die Transformation zum Datenaustausch mit zentralen Portalen oder die Integration der Daten in Datenbanken. Die Bearbeiter sind nicht von proprietären Datenbankmanagementsystemen abhängig, da alle Daten in Dateisystemen gespeichert werden können, falls sie nicht mit XML-Datenbanken verwaltet werden sollen. Die MEX-Werkzeuge sollen als Open Source Software zum Herunterladen angeboten werden, so dass sie individuellen Anforderungen angepasst und die Style Sheets hinzugefügt, verändert oder ausgetauscht werden können.

Das Projekt benötigt ungefähr ein Jahr Laufzeit und wird im Oktober 2007 abgeschlossen sein. Im April 2007 sollen erste Ergebnisse in Form einer CD den Teilnehmern der dritten europäischen Konferenz über EAD, EAC und METS zugänglich gemacht werden, die vom Bundesarchiv in Berlin während der deutschen EU-Präsidentschaft vom 24. bis 26. April 2007 unter dem Thema „Internationale Standards für digitales Archivgut“ (vgl. www.instada.eu) organisiert wird. Das Projekt soll im Rahmen eines Konferenzbeitrages und eines Informationsstandes vorgeführt und erläutert werden.

Arbeitsplan

Arbeitspaket 1:

Konzeptionierung der speziellen Elementauswahl, der entsprechenden Beschriftungen und der Anforderungen für Style Sheets der verschiedenen Werkzeuge. Mehrere Workshops des Projektteams mit den Programmierern sind erforderlich, um die

detaillierten Konzept für die Werkzeuge und ihre Definitionsdateien, die Beschriftungen und die Style Sheets zu erstellen. Die Personalkosten des Bundesarchivs werden aus dessen Etat bestritten. Die Fördermittel werden zur Finanzierung der Programmierungsleistungen der Firma eingesetzt, die die Software entwickelt und beratend zur Verfügung steht.

Arbeitspaket 2:

Anpassung und Programmierung der Software.

Dies wird von der gleichen Firma geleistet, die den MEX-Prototyp und die Pilotanwendung entwickelt hat. Die Programmierung umfasst das Erstellen neuer Definitionsdateien entsprechend den Schemas, die Anpassung der Arbeitsansichten und das Erstellen neuer Style Sheets, die an die Anforderungen der Standardtypen von Findbüchern angepasst sind. Dies erfordert einige Veränderungen im Vergleich zum Prototyp aufgrund neuer Entwicklungen von Eclipse und der Definition von XSLT 2.0.

Arbeitspaket 3:

Testen der entwickelten Software und Herstellen von Testszenarios.

Die verschiedenen Werkzeuge sollten in verschiedenen Arbeitsumgebungen getestet werden. Typische Szenarios werden zu diesem Zweck erstellt und die Testergebnisse dokumentiert. Dies soll geschehen, sobald Teile der Programmierarbeit erledigt sind, um Feedback für Verbesserungen zu erhalten. Diese Arbeit wird hauptsächlich von einer archivischen Beratungsfirma, erledigt.

Arbeitspaket 4:

Dokumentation der Entwicklungen und der Standardprofile für verschiedene Anwendungen.

Während der Arbeit wird die Dokumentation auf Grundlage der technischen Dokumentation der Programmierer und den Beobachtungen während der Tests erstellt. Dies dient u.a. zur Erstellung des Handbuchs, das den Benutzern zur Verfügung gestellt werden soll.

Arbeitspaket 5

Erstellen von Handbüchern für die Standardwerkzeuge.

Dies geschieht gleichfalls durch die Archivberatungsfirma zusammen mit dem Projektteam des Bundesarchivs.

Berlin, 3. Oktober 2006