

Tags: Archiv International, Bestand, Informatik,

Die Rechercheplattform ARGUS – Findmittel und digitalisierte Akten im Internet

ARGUS ist die Rechercheplattform des deutschen Bundesarchivs für Online-Beständeübersichten und Online-Findbücher. Ziel der Entwicklung von ARGUS war es, die Vorteile elektronischer Medien mit denen der früheren gedruckten Findbücher zu verbinden.

Seit 2005 sind die Erschliessungsangaben zu den Beständen der Stiftung «Archiv der Parteien und Massenorganisationen der DDR im Bundesarchiv» in ARGUS im Internet recherchierbar. Inzwischen wurde das Angebot auf andere Abteilungen des deutschen Bundesarchivs ausgedehnt. Ende 2011 war die Beständeübersicht des Archivs mit Kurzbeschreibungen zu mehr als 6000 Beständen, dazu knapp 2100 Online-Findbücher und darin 1,96 Mio. Verzeichnungseinheiten nutzbar. Darunter sind auch mehr als 700 retrokonvertierte ältere Findbücher. Auch ohne ausdrückliche Werbung wird ARGUS mit ständig steigenden Zahlen intensiv genutzt. Die Steigerungsrate von 100% pro Jahr seit 2005 war 2010 bei über 10 Mio. Seitenaufrufen angekommen. Im November 2011 wurden erstmals in einem Monat über 1,6 Mio. Seitenaufrufe gezählt. Ausserdem ist festzustellen, dass Anfragen an das Archiv häufig Rechercheergebnisse aus ARGUS zitieren. Was macht Argus für angemeldete wie neue Benutzer des Bundesarchivs offenbar attraktiv? Vielleicht hat es etwas damit zu tun, dass ARGUS bei der Präsentation von Findmitteln, bei der Darstellung von Digitalisaten wie bei der technischen Realisierung für die Nutzung im Internet einige neue Wege eingeschlagen hat. Sie sollen im Folgenden erläutert werden und dabei soll versucht werden, Lehren aus den bisherigen Erfahrungen zu ziehen.

Die Recherche in ARGUS

ARGUS ist ein Platz im Internet für die Recherche in der Online-Beständeübersicht und den Online-Findbüchern des Bundesarchivs. Es ist eine Rechercheplattform, in der die Informationen zum Archivgut übergreifend durchsucht, verglichen und genutzt werden können, um eine Nachforschung in den Archivalien vorzubereiten. ARGUS zeigt, was alles bereitsteht, und es vermittelt einen Eindruck davon, woher das Archivgut ursprünglich stammt. Es ist möglich, hier vor dem Archivbesuch bereits Materialien auszuwählen, die für eine Fragestellung relevant sein können. Dabei ist man zunächst nicht auf fremde Unterstützung angewiesen, kann aber präzise nachfragen, wenn es erforderlich sein sollte.

Ziel der Entwicklung von ARGUS war es, die Vorteile elektronischer Medien mit denen der früheren gedruckten Findbücher zusammenzubringen. Die Funktionen, die Bücher aufgrund ihrer Form besitzen, wie die äussere Abgrenzung von anderen Büchern samt deren Inhalt und damit der Bestände voneinander, die Einbettung jeder Einzelinformation über Einheiten des Archivguts in ihre Zusammenhänge, die Anzeige der aktuellen Gliederungsstelle im Seitenkopf, die fixierte Reihenfolge der Titel in einer Gliederungsgruppe oder der Überblick über das Findbuch im Inhaltsverzeichnis sind erhalten geblieben und mit den Möglichkeiten der Volltextsuche über alle Erschliessungsinformationen kombiniert worden. Findbücher sind mit ihren äusseren Grenzen in ARGUS deutlich erkennbar. Ausserdem ist ihre innere Struktur ständig präsent, und sie ist zudem zur Navigation nutzbar. Es kann gleichzeitig über alle Erschliessungsinformationen gesucht werden oder auf Wunsch auch zwischendurch mit einem neuen Begriff in einem einzelnen Findbuch. Schliesslich kann man in den Findbüchern in ARGUS blättern und wie im Buch einen von den erschliessenden Archivfachkräften erstellten Index nutzen.

Das Titelblatt mit Grundinformationen deutet bei der Navigation von Fundstelle zu Fundstelle den Wechsel zu einem anderen Bestand an. Die Einleitung gibt Erläuterungen zum Bestand. Das Inhaltsverzeichnis mit den Kapitelüberschriften sorgt für den Überblick. Die Verzeichnungseinheiten werden innerhalb der Gliederungsgruppen mit ihrem Titel samt einer eventuellen Erläuterung dazu im Enthältvermerk, mit ihrer Laufzeit und ihrer Signatur in der bei der Erschliessung festgelegten Reihenfolge aufgelistet. Layoutmerkmale werden eingesetzt, um wie in einem Buch als nonverbale Informationsträger zu wirken, aus denen Schlüsse über Zusammenhänge gezogen werden können. Denn eindeutige und stabile Kontexte sichern nicht zuletzt die Glaubwürdigkeit und Verlässlichkeit der Findmittel, da sie es erleichtern, eine wissenschaftliche Aussage argumentativ auf Archivgutfunde gründen zu können.

Gleichzeitig werden alle internen Referenzen im analogen Findbuch durch Links ersetzt. So wird das Inhaltsverzeichnis als ein anklickbarer Navigationsbaum genutzt, der dauerhaft am linken Rand präsent ist. Der Klick auf eine Überschrift öffnet daneben die entsprechende Gliederungsgruppe, die als Ganze durchgesehen werden kann. Ebenso öffnet sich der Index in einem überlagernden Fenster und der Klick auf die dort als Verweis angegebene Signatur öffnet das Findbuch darunter an der entsprechenden Stelle.

Mit der Volltextsuche kann jeder bei der Erschließung verwendete Begriff gefunden werden. Dabei können anhand der Struktur relevante Teile bereits vor der Suche ausgewählt werden, um die Menge der anschließend aufgelisteten Resultate selbst sinnvoll einzugrenzen. Nach der Suche bleibt die Navigation von Fundstelle zu Fundstelle auch bei einer zwischenzeitlichen Navigation durch die Strukturen oder einer eingebetteten Suche nach einem anderen Begriff in einem Findbuch verfügbar.

Das Gleiche gilt für die Beständeübersicht. Ihre kurzen Bestandsbeschreibungen integrieren auch diejenigen Bestände in die Recherche, zu denen noch keine Online-Findbücher vorliegen. Durch die Verknüpfung ihrer Bestandsbeschreibungen mit dem jeweils zugehörigen Findbuch wird eine Architektur mit mehreren Ebenen aufgebaut, die miteinander verknüpft sind. Die Beständeübersicht als oberste Ebene strukturiert die Gesamtheit der Online-Findbücher. Von den Bestandsbeschreibungen aus sind die Findbücher direkt zu öffnen. Von den Findbüchern aus kann man die bereits vorhandenen Digitalisate der Aktenseiten ansehen.

Die Kombination navigierender Recherche und übergreifender Suche nach Begriffen in dieser Präsentationsform macht frühere Diskussionen darüber, welcher von beiden Alternativen der Vorzug zu geben wäre, obsolet. ARGUS bietet beides zur selben Zeit und zum jederzeitigen Wechsel. Dabei werden neue Funktionen möglich, die sich erst aus der Kombination ergeben. Wird etwa beim genaueren Ansehen eines Bestandes festgestellt, dass in dem interessierenden Kontext bestimmte, bisher noch nicht bekannte Begriffe verwendet werden, kann ihre Nutzung in anderen Zusammenhängen oder anderen Beständen unmittelbar überprüft werden. Das ist zusätzlich eine so nicht geplante, aber nützliche Funktion für den Umgang mit fremden Sprachen, deren Spezialausdrücke trotz gewissen Kenntnissen noch unbekannt sind. So hilft die Präsentation ohne Zusatzaufwand beim Erlernen von spezifischen, in bestimmten Zusammenhängen genutzten Begriffen zusammen mit ihrem häufig kontextabhängigen Bedeutungsumfeld.

Digitalisate von schriftlichem Archivgut werden in ARGUS aus den Online-Findbüchern heraus über ein Kamerasymbol bei dem Titel der Akte mit einem speziell dafür entwickelten Präsentationsmodell einsehbar gemacht. Bis Ende 2011 konnten so 2 Mio. Seiten in 20 Online-Findbüchern bereitgestellt werden, darunter die vollständigen Bestände aus den Büros von Walter Ulbricht und Erich Honecker sowie die Druckgenehmigungsakten für Belletristik des Kulturministeriums der DDR.

Der technische Hintergrund

ARGUS ist ausgesprochen schnell. Kurz nach dem Start einer Suche über alle Angaben stehen die Suchergebnisse auf der Einstiegsseite in der Reihenfolge der Struktur zur Navigation in die Findbücher hinein bereit, von wo aus dann Fundstelle nach Fundstelle weiter angesehen werden kann. ARGUS nutzt die von der Archivschule Marburg entwickelte Suchmaschine Midosa-SEARCH, die mit Sets von statischen HTML-Seiten arbeitet. Alle Daten werden in dem internationalen EAD-Standard (encoded archival description) in ARGUS integriert. Aufbereitet und für die Installation bereitgestellt werden sie mit den Werkzeugen Midosa-XML oder MIDEX. Das betrifft die direkt hiermit erstellten Dateien ebenso wie exportierte Daten aus BASYS, der Archivdatenbank des Bundesarchivs.

Die in ARGUS eingesetzte Technik verwendet als konzeptionelle Grundeinheit eine Textdatei, nämlich das Findbuchdokument zur logischen Zusammenfassung aller Erschliessungsangaben und Informationen zu einem Bestand, das auch in einer Instanz als übergeordnete Beständeübersicht für das ganze Archiv funktionieren kann. Auch der Export aus der Erschliessungsdatenbank BASYS wird in Findbuchform gebracht und dann integriert. Bei der Aufbereitung für die Integration in ARGUS am archivischen Arbeitsplatz wird für jedes Findbuch automatisch ein Paket erstellt, das zum einen das Findbuchdokument und zum anderen einen daraus generierten Satz an HTML-codierten Dateien umfasst, die mithilfe von temporären Identifikatoren für die Präsentation miteinander verknüpft sind. Jedes Findbuch wird nach dem Hochladen auf den ARGUS-Server gegen das EAD-Format validiert und mit dem Volltextindexer Lucene indexiert. Es wird über den mitgelieferten Identifikator mit der zugehörigen Bestandsbeschreibung in der Beständeübersicht verknüpft und steht dann zur navigierenden Recherche wie zur übergreifenden Suche bereit.

EAD ist eine Form der XML-Codierung und wird hier in einem für diese Anwendung definierten Profil eingesetzt. Das Profil besteht aus einer Untermenge der Elemente und Attribute der Standarddefinition. Das betrifft besonders die Ebene der Gliederungsüberschriften, für die in Anlehnung an das Modell der Stufenerschliessung von ISAD(G) in EAD die Nutzung des vollen Umfangs aller für die Erschließung bereitgestellten Elemente erlaubt ist. Für die Erstellung der HTML-Ansichten wird ein Stylesheet eingesetzt, das an dieses Profil angepasst ist.

In EAD sind darüber hinaus zahlreiche Elemente verfügbar, die für die Verwaltung von Online-Findbüchern nützlich sind. So kann mit dem Element <eadid> jedem Findbuch eine maschinell lesbare Identifikation mitgegeben werden. In ihrem Kopfteil enthält eine EAD-Datei verschiedene Informationen zur bibliografischen Beschreibung des Findbuchs, die zu einem Titelblatt für das Findbuch und genauso zu einer Titelaufnahme für die Bibliothek zusammengestellt werden können. Die Datei hat einen Bereich für die Einleitung des Findbuchs wie Plätze für die Indizes, die Bearbeitende zu den Aktentiteln vergeben können. Mit speziellen, ineinander verschachtelbaren Elementen können die Gliederungsebenen mit Überschriften versehen werden und so die Strukturen abgebildet werden.

XML liefert die Voraussetzung für die doppelte Perspektive auf das Findbuch als ganzes Dokument wie auf seine inneren Strukturen. Es findet als Codierungsstandard zunehmend weitere Verwendung und wird inzwischen gern beim Austausch von Daten zwischen Datenbanken eingesetzt. In ARGUS erlaubt die Verwendung von XML die Orientierung auf die Findbücher als technische wie auch als archivfachliche Bearbeitungseinheiten. Die Komplexität der Arbeit mit der Installation wird dadurch im Vergleich zu einer Orientierung auf Verzeichnungseinheiten erheblich reduziert. Dieses Vorgehen erhöht die Übersichtlichkeit und reduziert die Aufwände für einzelne Arbeitsschritte, die für die Erstellung, Integration und Bereitstellung der Erschliessungsangaben erforderlich sind.

Die Einbindung der Digitalisate nutzt das gleiche Prinzip der Orientierung auf Dokumente, die jeweils einen logischen Komplex als Bearbeitungseinheit repräsentieren. Sie werden mit dem Standard METS in einer XML-Datei zu je einem intern strukturierten Set pro Aktentitel zusammengeführt und können dadurch bei der Einbindung in das EAD-Findbuch gemeinsam als je ein digitales archivisches Objekt behandelt werden.

Die Suchmöglichkeiten in ARGUS ebenso wie die Präsentation der Ergebnisse sind trotz der Dokumentenorientierung bei der Bearbeitung auf die Verzeichnungseinheiten bezogen. Das wird dadurch ermöglicht, dass sie als die Abfrageeinheiten (units of retrieval) definiert werden und dafür temporäre Identifikatoren bei der Umwandlung der Gliederungsgruppen nach HTML erhalten. Dieses Verfahren erhöht die Geschwindigkeit der Suche und der Anzeige der Treffer. Bei der Präsentation der Detailansicht nach einer Suche steht die erste Zeile der jeweiligen Verzeichnungseinheit mit entsprechender gelber Hervorhebung des Treffers in ihrem Text am oberen Rand der Anzeige innerhalb der vollständigen Gliederungsgruppe und inmitten der nicht gesuchten Titel. Die Detailansicht führt also zur Fundstelle des Titels, nicht nur zur Lokalisation des Treffers im Volltext.

Die Verwendung von XML in der Form von EAD hat positive Wirkungen auf die Arbeitsökonomie bei der Ergänzung des Angebots und bei der Recherche. Sie hat sich ebenso für die nationale und internationale Kooperation bewährt. Für das geplante Archivportal D wie auch für das europäische Archivportal ist jeweils ein sehr ähnliches EAD-Profil entwickelt worden. Die Bereitstellung von Online-Findbüchern, die in ARGUS integriert sind, für diese gemeinsamen Zugriffspunkte erfordert im Routinebetrieb minimalen Aufwand und kann mit einer automatischen Übersetzung des Datenformats von einem EAD-Profil in ein anderes geschehen. Die Verknüpfungen der Findbücher mit den METS-Dateien führt von den Portalen aus auf den Server beim Bundesarchiv, auf den auch ARGUS zugreift.

Die Bearbeitung der Daten für ARGUS

Zur Bearbeitung der Findbücher wie zur Integration von Digitalisaten wird das Werkzeug MIDEX verwendet. Es geht zurück auf eine von der Mellon-Stiftung 2005 bis 2007 geförderte Entwicklung, die das Werkzeug Midosa-XML um neue Funktionalitäten erweiterte. Damals konnte das Präsentationsmodell für Digitalisate entwickelt werden. Das Werkzeug wurde später noch mehr erweitert und erhielt zusätzliche Konvertierungsfunktionen beim Aufbau des Netzwerks SED-/FDGB-Archivgut zusammen mit den 13 Staatsarchiven der neuen Bundesländer.

MIDEX arbeitet ähnlich wie eine Textverarbeitung. Die Daten werden in Dateisystemen verwaltet, mit MIDEX geöffnet, bearbeitet, wieder abgespeichert oder in ein anderes Format konvertiert. Die Dateien können unabhängig von der Software kopiert und weitergegeben werden. Sie können ebenso mit jedem Texteditor oder XML-Editor geöffnet, gelesen und verändert werden und bleiben dabei voll funktionsfähig, solange die Grundsätze der XML-Syntax beachtet werden. Gleichzeitig steht im Werkzeug eine Vorschaufunktion bereit, sodass jederzeit das Ergebnis der Arbeit kontrolliert werden kann. Dadurch ist es ideal für eine dezentrale Bearbeitung auch ohne Netzanbindung und mit minimaler Ausstattung.

Für die Integration neuer Daten in ARGUS wurde eine Website entwickelt, auf der dezentral das Hochladen von Findbüchern, ihre Indexierung und Integration sowie auch das Löschen vorhandener Daten angestoßen und kontrolliert werden können. Das Ergebnis wird zunächst auf einem Testserver angezeigt, der in regelmässigen Abständen auf den Produktivserver kopiert wird.

ARGUS – eine neue Form der Findbuchpublikation

ARGUS ist eine Art Publikationsmedium geworden, mit dem archivische Erschliessungsinformationen mit wenig Aufwand und zudem schnell aktualisierbar im Internet präsentiert werden können. Es hat sich, wie an den Nutzungszahlen zu erkennen, als Rechercheplattform bewährt und offensichtlich ist die kritische Masse erreicht, die es nützlich macht. Es hat auf der anderen Seite Konzepte als operational erwiesen, die mit der Orientierung aller Bearbeitungsschritte auf archivische Komplexe als digitale Objekte, seien es die Findbuchdokumente, seien es die Sets von Digitalisaten einer Akte, ein grosses Rationalisierungspotenzial bei gleichzeitiger Wahrung hoher fachlicher Qualität eröffnen. Mit ARGUS ist ein Weg beschritten, der neue Möglichkeiten digitaler Publikation schafft und gleichzeitig nützliche Funktionen der analogen Welt erhält.

Eine Auswertung von Archivgut für neue Fragestellungen ist in jedem Fall mit Arbeit verbunden. Dafür belohnt sie jeden, der sich darauf einlässt, mit neuem Wissen, das noch niemand anderes kennt. Den Weg dazu versucht ARGUS zu erleichtern und von unnötigen Hürden zu befreien.

Links/Webseiten: www.argus.bundesarchiv.de www.archivgut-online.de

Netzwerk SED-/FDGB-Archivgut: www.bundesarchiv.de/sed-fdgb-netzwerk/

Europäisches Archivportal: www.archivesportaleurope.eu www.apenet.eu



Angelika Menne-Haritz

Deutsches Bundesarchiv

Abstract

Français

ARGUS est la plateforme de recherche des Archives fédérales allemandes pour l'accès en ligne aux inventaires des fonds et aux instruments de recherche numérisés.

Le développement de l'ARGUS répond au besoin de combiner les avantages des médias électroniques aux outils de recherche précédemment fournis sous forme imprimée. Pour ce faire, de nombreuses fonctions jusque-là gérées par les instruments de recherche ont été transférées sur la plateforme numérique. A titre d'exemple, on citera la page de couverture des inventaires accompagnée d'informations de base sur l'instrument de recherche, une introduction définissant l'inventaire, une table des matières, un index ou des annexes. La combinaison des modes de recherche et d'une approche plein texte dans ARGUS rend oiseuses toutes les discussions touchant la préférence à donner à l'une ou l'autre des approches vu qu'ARGUS offre toutes les possibilités en tout temps et pour toute requête.

ARGUS utilise le moteur de recherche MidosaSEARCH développé par l'Ecole de Marburg, lequel exploite des ensembles de pages HTML statiques. De plus, toutes les données seront intégrées selon le standard international EAD. Elles sont préparées et configurées pour cette normalisation, en recourant aux outils MidosaXML et MIDEX. Pour chaque instrument de recherche, un fichier XML est créé dont est tiré un ensemble de fichiers HTML. Les fichiers XML sont validés par rapport au format EAD et indexés à l'aide du moteur plein texte Lucene. Malgré l'accent mis dans le traitement sur les structures complexes (à savoir, les instruments de recherche), les moyens de recherche et la présentation des résultats sont fondés sur les unités de description, envisagées comme unité de base des requêtes (units of retrieval). L'utilisation de XML selon le schéma EAD est avantageuse non seulement en termes d'économie du travail, mais aussi dans le cadre des collaborations nationales et internationales vu que le Portail des archives en Allemagne comme celui développé dans le cadre européen feront appel à un profil très proche de l'EAD.

Comme le nombre d'utilisateurs en hausse en témoigne, l'ARGUS s'est avéré être une plateforme de recherche adéquate et qui dispose de la taille critique suffisante pour en faire un outil utile.