

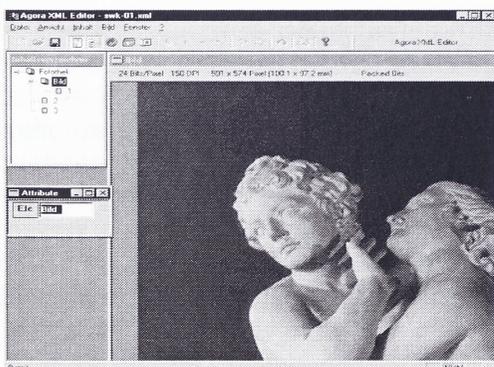
DIE STIFTUNG WEIMARER KLASSIK UND AGORA - KONZEPTION EINER EINHEITLICHEN LÖSUNG FÜR FOTOTHEK; BIBLIOTHEK, ARCHIV UND MUSEUM

Martin Fock-Althaus
Satz-Rechen-Zentrum Berlin
Bessemerstraße 83-91
12103 Berlin
Tel.: 030-75301-386, Fax 030-75301-319
e-mail: M.Fock-Althaus@srz-berlin.de

Das SRZ, gegründet 1969, beschäftigt sich seit seines Bestehens mit der Bearbeitung von elektronischen Datenbeständen. Standen in den Anfangsjahren noch Datenpflege im Satzbereich und maschinelle Datenkonversion im Vordergrund, so kamen Mitte der 80er Jahren Digitalisierungs- und Indizierarbeiten großer Datensammlungen (Patentschriften des deutschen und europäischen Patentamtes) hinzu. Diese Arbeiten entwickelten sich in der Folge zu einem eigenständigen Geschäftsbereich, der sich zunehmend auch mit elektronischem Archivieren, CD-ROM-Publikation und Printing on Demand befaßte. Mit dem Aufkommen des Internets wurden die vorhandenen elektronischen Publikationsschienen auf dieses Medium erweitert. Mit entsprechender Aufrüstung des Rechenzentrums (leistungsstarke Server und Datenleitungen, usw.) wurde eine Umgebung geschaffen, die es ermöglichte, Internet-Dienstleistungen im Hause anzubieten.

Im Zusammenhang vor allem mit Digitalisierungsarbeiten für verschiedene Bibliotheken wurde immer deutlicher, daß der Weg vom reinen Image (Monographie, Musikdruck, Graphik etc.) zum Dokument im Internet einer Automatisierung bedurfte. Einzellösungen für bestimmte Dokumentensammlungen waren nur sehr arbeitsaufwendig und zeitintensiv herzustellen und konnten absehbar die aufkommende Datenflut nicht bewältigen.

Ein erster Schritt ging zunächst in Richtung ‚Strukturierung von digitalisierten Dokumenten‘ Hier wurden benutzerfreundliche Werkzeuge geschaffen um webtaugliche Metadaten für die gescannten Datenströme einzelner Dokumente erstellen zu können. Die Standards ‚HTML‘, ‚SGML‘ und ‚XML‘ wurden immer berücksichtigt; keine Rolle allerdings sollte es spielen, um welche Art von Dokumenten es sich handelt. Es sollte möglich sein, auch vollkommen disparates Material, in verschiedensten Kollektionen zu erschließen und dieses später in einer einheitlichen Internetpräsentation darzustellen. Photographie und historischer Druck, Notenblatt und Reisebericht, hochwertiger Farbscan und einfaches Schwarz/Weiß-Image sollten über **eine** Benutzeroberfläche zu bearbeiten sein und in **einem** Format abgelegt werden können. Als übergreifendes Datenformat wurde schließlich XML gewählt, da dieses sich als Internet-Standardformat etabliert hatte.



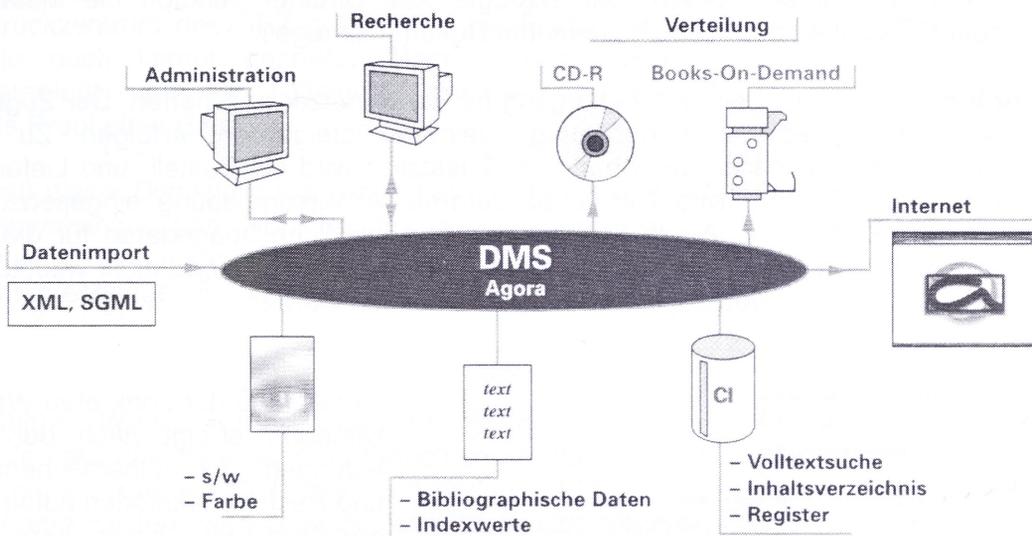
Erfassungsmaske des Agora XML Editor

In Entwicklungspartnerschaft mit dem Göttinger Digitalisierungszentrum (GDZ) wurde 1999 das Agora System konzipiert, das den gesamten Komplex ‚digitale Kollektion – Internet-Präsentation‘ abdecken sollte. Sowohl was die Administration der digitalen Sammlungen anbelangt als auch was hinsichtlich der Generierung von HTML-Daten und Erzeugung von webfähigen Images notwendig ist, sollte direkt aus dem System steuerbar sein. Der Systemverwalter sollte die Möglichkeit haben, über den Import von XML-Dokumenten automatisch alle weiteren Komponenten für die Web-Darstellung aufbauen zu können. Konkret heißt das: wenn ein Dokument in der Agora Administration vorliegt, kann es auch im Internet ohne weiteres recherchiert und angezeigt werden. Navigationsmöglichkeiten innerhalb eines digitalen Dokuments sind immer gegeben, da für jedes Dokument ein (virtuelles) Inhaltsverzeichnis aufgebaut wird, das den Dokumentaufbau hierarchisch anzeigt und alle Links zu den gewünschten Elementen und Dokumentseiten enthält. Der Aufruf der vom Benutzer ausgewählten Dokumentseiten erfolgt dynamisch über ein Java-Servlet.

Ebenfalls Standard ist die Generierung von bibliographischen Seiten zu jedem Dokument, je nach Inhalt der vorher in XML erschlossenen Metadaten.

Das System ist grundsätzlich frei konfigurierbar, d.h. das Design von Web-Seiten und Recherchemasken kann vom Benutzer (über HTML-Templates) nach seinen Vorstellungen gesteuert werden, ebenso sind z.B. Recherchebegriffe beliebig formulierbar und kombinierbar. Es kann sowohl innerhalb der erschlossenen Metadaten als auch im Volltext recherchiert werden.

Ohne Abhängigkeiten stellt sich das System auch in Hinsicht auf Schnittstellen und Hardwareumgebung dar. Ein im Grunde beliebiges Datenbanksystem kann eingesetzt werden, es muß lediglich SQL-fähig sein und über eine ODBC-Schnittstelle erreicht werden können, die Systemplattform kann sowohl Unix/Linux als auch Windows NT heißen.



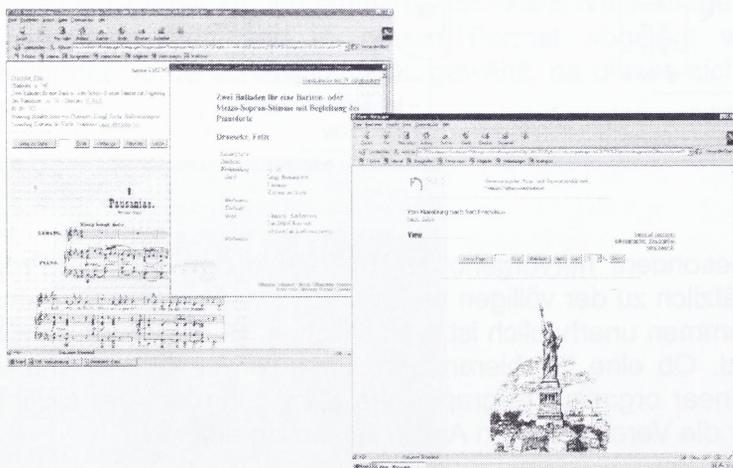
Ein Aspekt aber sei noch einmal besonders hervorgehoben, da er für die folgende Projektbeschreibung von Bedeutung ist: zusätzlich zu der völligen technischen Offenheit des Systems ist entscheidend, daß es für Agora vollkommen unerheblich ist, aus welchem Bereich die Digitalisate kommen und wie sie strukturiert sind. Ob eine tief hierarchisch ausgeformte wissenschaftliche Arbeit im Volltext vorliegt oder eine linear organisierte graphische Sammlung, bei der ein Image einem Dokument entspricht, macht für die Verarbeitung in Agora keinen Unterschied.

In einem aktuellen Projekt, welches das SRZ mit der Stiftung Weimarer Klassik durchführt, können die Vorteile des offenen Agora Systems verdeutlicht werden.

Die Fotothek der Stiftung Weimarer Klassik bearbeitet alle Anfragen (wissenschaftlicher und kommerzieller Art) nach Bildvorlagen aus den Sammlungen des Goethe-Nationalmuseums, der Herzogin Anna Amalia Bibliothek und des Goethe- und Schiller-Archivs. Der gegenwärtig vorhandene Bestand umfaßt ca. 120.000 Bildmotive, wobei der jährliche Zuwachs etwa 5.000 Einheiten beträgt. Neben der Bearbeitung von Kundenanfragen, mit den damit einhergehenden Aufgaben, wie Beratung, Prüfung von Reproduktionsfähigkeit einzelner Dokumente, Kontrolle, Rechnungsstellung und Versand, liegen die weiteren Aufgaben der Fotothek in der Archivierung, Erschließung und Erhaltung des Bildarchivs. Alle diese Arbeiten werden bislang weitgehend ohne elektronische Unterstützung durchgeführt.

Für die Umsetzung dieser Arbeitsabläufe in ein elektronisches Dokumentenmanagement- und Archivierungssystem wurden folgende Schritte festgelegt:

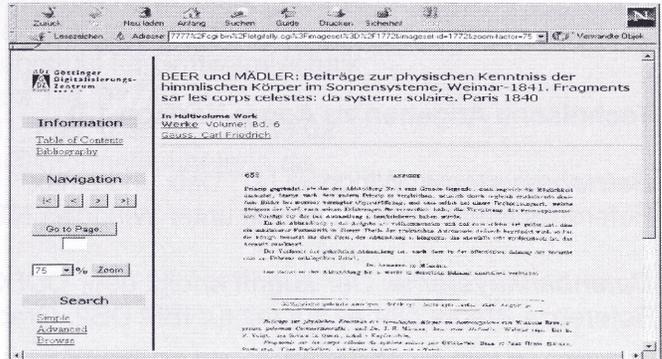
1. **Digitalisierung:** Der Bestand an Bildvorlagen ist zu digitalisieren, für ausgewählte Dokumente werden Digitalisate von hoher Qualität erzeugt, so daß eine auch drucktechnisch hochwertige Reproduktion erstellt werden kann. Für Teile der Sammlung genügt eine reduzierte Imagequalität, jedoch soll diese besser sein, als es die einfache Web-Darstellung erfordert. Schon vorhandene gescannte Dokumente sollen in das System übernommen werden können. Mit dem Agora Konverter können die Scans – falls erforderlich – in ein webtaugliches Bildformat (GIF oder JPEG) konvertiert werden.
2. **Erschließung, Übernahme von Metadaten:** Das vorhandene Karteikartensystem wird nach XML übertragen. Das Werkzeug hierfür ist der Agora XML Editor. Es wird ein Datenmodell zugrundegelegt, das die Katalogkartenstruktur nachvollzieht, jederzeit aber für Erweiterungen und Änderungen offen ist. Vorhandene Erschließungsdaten (etwa aus dem PICA-Katalog) können direkt übernommen werden. Die erzeugte XML-Struktur verknüpft die Metadateninformation der Katalogkarten mit den gescannten Dokumentimages.
3. **Internet-Anbindung:** Es wird ein Internetzugang für die Benutzer geschaffen. Der Zugriff soll auf die Bilddaten mit reduzierter Auflösung (kleinere Dateigrößen) erfolgen. Zu jedem Dokument soll ein Metadatensatz abrufbar sein. Zusätzlich wird ein Bestell- und Liefersdienst integriert. Das Agora System wird hierfür als zentrale Arbeitsumgebung eingesetzt. Beim Import der XML-Dokumente in Agora können sowohl zusätzliche Imagedaten für die Web-Darstellung erzeugt (sofern nicht schon vorhanden) als auch bestehende nicht webtaugliche Bildformate in entsprechend webtaugliche Formen konvertiert werden.



Der Aufbau der konkreten Web-Anwendung erfolgt nach den Festlegungen für Oberflächendesign und Recherchekriterien automatisch aus dem Agora System heraus. Der Benutzer kann sich, nachdem das System die HTML-Dokumentseiten generiert hat (ein einmaliger Vorgang für jedes Dokument), über die Eingabe verschiedener Suchkriterien oder Textstrings zur Volltextrecherche das gewünschte Dokument auf den Bildschirm holen und er kann alle bibliographischen Zusatzinformationen zu seinem

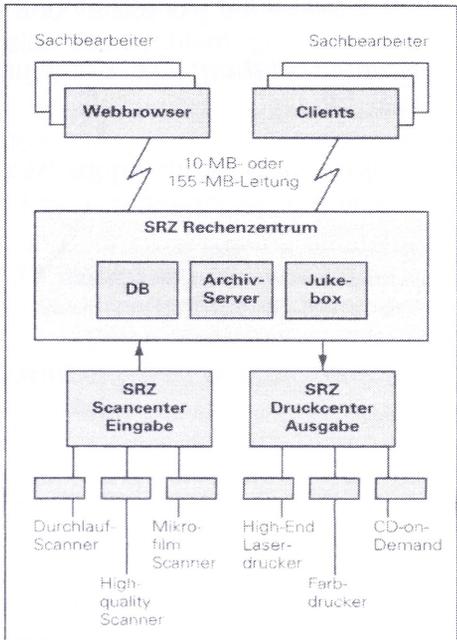
Dokument abrufen. Die Navigation im Dokument ist über einen Link zum Inhaltsverzeichnis möglich oder über das klassische Blättern im Dokument, verbunden mit einer GoTo-Funktion, die direkt die gewünschte Seitennummer aufruft und zwischen logischer und physikalischer Paginierung unterscheiden kann.

Liegt das aufgerufene Dokument in textcodierter Form vor, kann der Benutzer das Dokument auch mit Hilfe des ‚Highlighting on Facsimile‘-Verfahrens untersuchen, d.h. er kann sich Textelemente direkt auf der Faksimile-Seite anzeigen lassen.



Schließlich kann er auch eine elektronische Bestellung aufgeben. Diese löst im System eine Lieferaktion aus, die Zugriff auf die Bilddaten hoher Qualitätsstufe hat. Der Besteller erhält den gewünschten Datensatz in der *Originalqualität* als Digitalisat (per e-mail, per CD, etc.). Dieses kann in Form reiner Imagedaten; in Form einer mit XML-verknüpfte Bildsammlung oder als ein mit Bookmarks strukturiertes PDF-Dokument geschehen. Geplant ist zusätzlich, die Möglichkeit zu schaffen, via Internet Druckaufträge vergeben zu können. Diese Aufträge können in elektronischer Form an das angebundene Druckzentrum des SRZ gehen, das die Produktion eines wie auch immer gearteten Reprints des Dokuments veranlaßt - z.B. als Digitaldruck einer Monographie oder als Proof einer Graphik.

Daß dieser Dienstleistungsbereich auf ein funktionierendes Verwaltungs- und Abrechnungssystem angewiesen ist, leuchtet ein. Agora wird daher in diesem Projekt um die entsprechenden Funktionalitäten erweitert.



Die Stiftung Weimarer Klassik besteht nicht nur aus der erwähnten Fotothek, vielmehr ergänzen sich die verschiedenen anderen Einrichtungen (Goethe-Schiller-Archiv, Goethe-Nationalmuseum, Herzogin Anna Amalia Bibliothek) und sind in vielfältiger Weise miteinander verbunden. Das Agora System wird zunächst eine Verbindung der verschiedenen Nachweissysteme dieser Einrichtungen herstellen: alle hier vorhandenen bibliographischen Informationen zu den Dokumenten der Fotothek werden über den Internetzugriff ebenfalls verfügbar sein.

Wenn in der Zukunft eine der Institutionen der Stiftung sich dafür entscheidet, eine weitere Dokumentensammlung im Internet zu präsentieren, so kann diese Sammlung ohne Aufwand als neue Agora Kollektion angelegt und bearbeitet werden. Sie wird einfach den bestehenden digitalen Sammlungen zur Seite gestellt. Ein Systemkonflikt wird so grundsätzlich vermieden, das Wachsen der digitalen Bibliothek kann sich an inhaltlichen Aspekten orientieren und muß nicht auf system- oder softwarespezifische Gegebenheiten Rücksicht nehmen. Der langfristige Zugriff auf

sämtliche elektronische Daten kann sowohl über interne Datenspeicher gewährleistet werden als auch durch Datenvorhaltung auf externen Servern im SRZ. Erweiterungsmöglichkeiten sind allen Richtungen gegeben, Begrenzungen bestehen weder im technischen noch inhaltlich dokumentspezifischen Bereich.

Technische Angaben zu Agora Version 1.0

Betriebssysteme: Windows NT, Unix und Linux

Referenzsysteme gibt es zurzeit unter Windows NT 4.0 und Sun Solaris

Datenbanksysteme: Der Zugriff erfolgt über ODBC bzw. JDBC

Referenzsysteme gibt es zurzeit für IBM DB2, Oracle, Sybase, SQL-Server und MS-Access

Client-Server-Architektur: Der Agora-Server ist als Servlet in Java programmiert und bietet Plattformunabhängigkeit. Jeder Internetbrowser ist als Client geeignet. Der Agora Administrationsclient läuft unter Windows NT. Agora ist multiuserfähig

Kommunikation: Die Kommunikation mit dem Agora-System erfolgt über Standardprotokolle (z.B. http).

Konfigurierbarkeit: Agora ist sowohl bezüglich der Datenstruktur als auch bezüglich der Web-Präsentation sehr anpassungsfähig

Recherche: Die Recherche in Agora ist über Meta- und Volltextdaten möglich. Agora setzt den Verity Informationsserver als leistungsfähiges Fremdprodukt ein.

Referenzen und Beispielanwendungen: <http://www.agora.de/samples.html>