

Datenhaltungsstandards und Möglichkeiten ihrer Anwendung bzw. Aufbereitung

(mit Beispielen aus dem *Edirom*-Projekt)

Daniel Röwenstrunk

23. Dezember 2006

Ein Vortrag¹ für das Kolloquium "Digitale Medien und Musikedition" vom 16. bis zum 18. November 2006 von der Akademie der Wissenschaften in Mainz.

I.

Ich möchte mich im ersten Teil dieses Vortrages bewusst an all diejenigen wenden, die bis jetzt noch nicht mit elektronischer Datenhaltung, etwa zum Zweck digitaler Musikedition in Kontakt gekommen sind, um auch Ihnen die Chance zu geben, die bisher gehörten und noch folgenden Vorträge besser zu verstehen. In einem zweiten Teil soll es dann um unsere, d.h. neben meiner auch um Herrn Keppers und Herrn Veits, Vorstellung einer digitalen Musikedition gehen.

Prinzipiell versteht man unter der elektronischen Datenhaltung zunächst nur die Übersetzung von realen Gegenständen oder Informationen in eine von Maschinen lesbare Form. Das bedeutet noch nicht, dass der Computer auch versteht, was er speichert. Eine Zeitschriftenseite könnte zum Beispiel als Bild oder als Transkription in ein elektronisches Datenformat überführt werden. Sie sehen daran aber sehr schnell wesentliche Unterschiede der Nutzbarkeit der Informationen in Abhängigkeit von den eingesetzten

¹Der Text entspricht bis auf wenige Angleichungen und Erweiterungen (z.B. der Beispiele) dem mündlichen Vortrag während der Tagung.

Technologien. Sie können von einem Bild zwar in der Regel bessere Ergebnisse hinsichtlich einer Reproduktion eines Originals, von einer Transkription hingegen bessere Chancen zur weiteren Verwendung der Inhalte erwarten. Wenn es also heißt, dass elektronische Datenhaltung prinzipiell wesentliche Vorteile gegenüber jeglicher analoger Archivierung oder Aufbereitung von Informationen bietet, dann ist das nicht immer richtig und wesentlich davon abhängig, in welche digitale Form die Inhalte überführt werden sollen. Es ist durchaus korrekt, dass Sie deutliche Vorteile hinsichtlich der weiteren Nutzbarkeit erzielen können, wenn Sie eine Transkription der Zeitschriften-seite speichern, da sie auf diese Art zum Beispiel durchsuchbar wird und die Möglichkeit einer inhaltlichen Aufbereitung bietet. Aber auch hier spielt das Datenformat, in dem Sie das Dokument hinterlegen, eine wesentliche Rolle für die tatsächliche Verwertbarkeit der Informationen.

Ich möchte nun für diejenigen unter Ihnen, die technisch weniger versiert sind, das bereits viel zitierte XML in bewusst nicht-technischer Weise von Grund auf erläutern. Die eXtensible Markup Language (XML) ist eigentlich kein wirkliches Datenformat, sondern, wie der Name schon sagt, eine erweiterbare Auszeichnungssprache, mit derer Hilfe sowohl Datenformate definiert als auch Inhalte in so festgelegten Strukturen abgelegt werden können. Unter einem Datenformat versteht man die Definition der Art und Weise, in der die Inhalte eines Formates codiert werden müssen und folglich verstanden werden können; für Dateien, die mit XML ausgezeichnet werden sollen, werden diese Formate mit sogenannten Document Type Definitions (DTD) oder XMLSchemata beschrieben. XML ist also ein Werkzeug, mit dem sowohl Formate erstellt,² als auch die Inhalte nach den Regeln dieser Formate strukturiert werden können. Es stellt dabei keinerlei Voraussetzungen an die Art der Inhalte oder die Strukturierung derselben, es setzt lediglich ein paar Regeln für den Aufbau einer XML-Datei fest. XML-Dateien sind hierarchisch strukturierte Daten, die in frei wählbarer Form vorliegen.

In Abbildung 1 sehen Sie eine XML-Datei, die ein sehr stark vereinfachtes Objekt vom Typ "Buch" beinhaltet. Das Buch wird durch die Unterelemente Titel des Buches, Autor, ein Verlag und ein Erscheinungsdatum repräsentiert. Sie sehen in diesem sehr kurzen Beispiel schon alle wesentlichen Bestandteile einer XML-Datei; es gibt Elemente, wie z.B. das "book"-Element oder das "title"-Element, es gibt Attribute, wie in diesem Beispiel das Attribut "value" innerhalb des "publishingDate"-Elements und es gibt den Inhalt dieser Elemente und Attribute, der zumindest bei den meisten Elementen

²XMLSchemata sind im Gegensatz zu DTDs selbst auch XML-Dateien, die in ihrer Struktur wiederum durch Schemata definiert werden.

```
<book>
  <title>Das Universum in der Nussschale</title>
  <author>Stephen W. Hawking</author>
  <publisher>dtv</publisher>
  <publishingDate value="2004-01-" />
</book>
```

Abbildung 1: Beispiel einer Buch-Repräsentation in XML

wiederum aus Elementen bestehen könnte³. Elemente werden durch ein in spitze Klammern gefasstes Schlüsselwort für das Objekt geöffnet (`<title>`); geschlossen werden sie durch das gleiche Konstrukt aus Klammern und Schlüsselwort mit einem zusätzlichen Schrägstrich nach der öffnenden spitzen Klammer (`</title>`). Diese Konstrukte werden "tag" bzw. in der Mehrzahl "tags" genannt. Alle Informationen zwischen diesen Tags werden als Inhalt des Elementes aufgefasst. Ein leeres Element wird entweder durch den fehlenden Inhalt zwischen den tags oder durch eine besondere Konstruktion angezeigt, welche sich aus einem öffnenden tag mit zusätzlichem Schrägstrich vor der schließenden spitzen Klammer (`<publishingDate/>`) ergibt. Attribute sind Paare aus Schlüsselwort und Wert und können außer Text keine weiteren Elemente enthalten. Sie bieten sich somit für Informationen an, die nicht weiter aufgeschlüsselt werden sollen, in Abbildung 1 das Erscheinungsdatum. Attribute werden immer innerhalb eines öffnenden tags eines Elementes platziert, wodurch eine feste Bindung des Attributes an ein Element stattfindet; sie werden durch ein Leerzeichen getrennt hinter das Schlüsselwort des tags in der Form `Name="Wert"` geschrieben. Die Schlüsselwörter der Elemente und Attribute sind frei wählbar,⁴ sie sollten den Inhalt aber bestmöglich beschreiben, sodass die XML-Datei ohne weitere Hilfestellungen zu der Bedeutung der Auszeichnung lesbar wird. Zumindest für Inhalte von internationalem Interesse sollten Element- und Attributnamen in englischer Sprache verwendet werden. Wenn also möglichst gut beschreibende Namen für die Elemente gewählt werden, beschreibt sich der Inhalt einer XML-Datei selbst und bietet somit erhebliche Vorteile gegenüber anderen Formaten.

Ein etwas umfassenderes Beispiel für eine sich selbst erklärende XML-Datei sehen Sie in Abbildung 2; an dem alles umschließenden Element "document"

³Ob ein Element weitere Elemente enthalten kann, hängt von der zugrunde liegenden Struktur (DTD oder XMLSchema) des Dokumentes ab.

⁴Die einzigen Einschränkungen beziehen sich auf einige wenige Sonderzeichen und das Verbot, eine Zahl als erstes Zeichen zu nutzen.

```

<document>
  <meta>
    <source>Abend-Zeitung, Jg. 1, Nr. 238 (04. Oktober
1817)</source>
  </meta>
  <content><headline1>Nachrichten aus dem Gebiete der Künste
und Wissenschaften.</headline1><headline2>Chronik der Königl.
Schaubühne zu Dresden.</headline2>
    <sourcePage>[2v/a]</sourcePage><p>Den 25. September. Im
Königl. Hoftheater. <x>Die vornehmen Wirthe</x>. Komische Oper
in 3 Akten, nach dem Französischen von <x>I. R. von
Seyfried</x>, Musik von Catèl.</p><p>Eine nach allen ihren
Theilen höchst angenehme Production, welche uns jedoch hinter
dem Schimmer fließender Leichtigkeit manche für Instrumente
sehr schwierige Stellen zu haben schien, wie z. b. in der
Ouvertüre einige Gänge für die Violine; in dem lieblichen
Vorspiel zum dritten Akt, das Flötensolo, für dessen
vortreffliche Ausführung Herr Steudel mit dem lautesten
Beifall wiederholt belohnt wurde. Ueber die Musik selbst kein
Wort, da eine gewichtige Feder in diesen Blättern schon das
Gehörige darüber gesagt hat.</p></content>
</document>

```

Abbildung 2: Beispiel eines Zeitschriftenartikels in XML

können Sie erkennen, dass es sich um ein Dokument im weitesten Sinne handelt. Bei näherer Betrachtung des Inhalts des Dokumentes fällt zunächst der "meta"-Bereich auf, der weitere Aufschlüsse über das Dokument gibt, den Angaben im "source"-Element zufolge liegt Inhalt aus der Abend-Zeitung Jahrgang 1 aus dem Jahr 1817 vor.⁵ Der Bereich, der durch die *tags* des "content"-Bereiches aufgespannt wird, enthält offensichtlich den eigentlichen Inhalt des Dokumentes, der zusätzlich mit rein auf die Darstellung bezogenen Auszeichnungen wie kursiver (<i>) oder gesperrter (<x>) Schrift und mit einfachen strukturellen Auszeichnungen wie den Überschriften (<headline1> bzw. <headline2>) versehen ist. Es liegt also ein Auszeichnungsgrad vor, der sich auf die Reproduktion des Originals beschränkt.

⁵Aus Gründen der Übersichtlichkeit sind die Metaangaben nur verkürzt dargestellt. Es handelt sich um einen Auszug eines Artikels aus der Dresdner Abend-Zeitung, Jahrgang 1, Nr. 238 vom 04. Oktober 1817 auf der Seite 2 verso, Spalte 1.

Die Beispiele der XML-Inhalte und somit alle anderen XML-Dateien auch sind durch die Einfachheit des Dateiformates "Textdatei" in Kombination mit der selbst erläuternden Struktur tatsächlich langfristig haltbar. Es wird mit großer Sicherheit auch in einem für elektronische Datenhaltung schon sehr langem Zeitrahmen von 50 Jahren noch Programme geben, die in der Lage sein werden, XML-Dateien wenigstens in dieser nicht graphisch aufbereiteten Ansicht zu öffnen. Programme, die diese XML-Dateien interpretieren, auswerten und in graphische Oberflächen integrieren können, müssen nicht unbedingt so lange Lebensdauern aufweisen, Sie werden also unter Umständen die Inhalte selbst interpretieren müssen. Es gibt noch einige weitere Vorteile von XML-codierten Daten gegenüber anderen Formatierungsmöglichkeiten für die Archivierung, die an dieser Stelle nicht weiter angesprochen werden sollen.⁶ Vielleicht können Sie mit diesem kurzen Einblick in die einfache Struktur von XML-Dateien und dem Wissen, dass bei fast allen großen Digitalisierungsprojekten XML eingesetzt wird, auch für sich die Vorteile dieser Auszeichnungssprache erkennen.

Um die Möglichkeiten von XML und die sich daraus ergebenden Vorteile noch stärker zu verdeutlichen, möchte ich Ihnen nun die Potenziale der Sprache bezüglich der Nutzung der Daten weiter ausbreiten. Wie wir anhand der Beispiele (Abbildungen 1 und 2) gesehen haben, können beliebige Inhalte in XML aufbereitet und gespeichert werden; so ist es möglich, die Inhalte noch über die bloße Reproduktion des Originals hinaus näher zu beschreiben. Im Beispiel in Abbildung 2 lässt sich aus dem Text erkennen,⁷ dass ein Stück besprochen wird, welches "Die vornehmen Wirthe" heißt. Es wäre also möglich, das Werk als solches innerhalb der XML-Datei kenntlich zu machen, zudem unter Umständen noch den Autor, den Komponisten oder etwa das Erscheinungsdatum anzugeben. Auch die im Text auftretenden Personen⁸ könnten als solche markiert und z.B. mittels einer Identifikationsnummer mit einer anderen Personen-XML-Datei mit weiterführenden Informationen verknüpft werden. Nach diesem Schema ließe sich der gesamte Inhalt des Dokumentes beschreiben, was offensichtlich schnell komplexe Formen annehmen kann. Da es sich bei dem in Abbildung 2 dargestellten Beispiel um einen Auszug aus dem originalen Dokument handelt, möchte ich Ihnen

⁶Vgl. hierzu z.B. Auszug "Using XML" aus Visual Basic® .NET Developer's Guide to ASP .NET, XML and ADO.NET von Chris Kinsman und Jeffrey McManus <http://www.awprofessional.com/articles/printerfriendly.asp?p=25855&rl=1> (Einsichtsdatum: 23.12.2006).

⁷Vgl. Auszug aus dem Dokument: Den 25. September. Im Königl. Hoftheater. <x>Die vornehmen Wirthe</x>. Komische Oper in 3 Akten...

⁸Vgl. Auszug aus Abbildung 2: Komische Oper in 3 Akten, nach dem Französischen von <x>I. R. von Seyfried</x>, Musik von Catél.

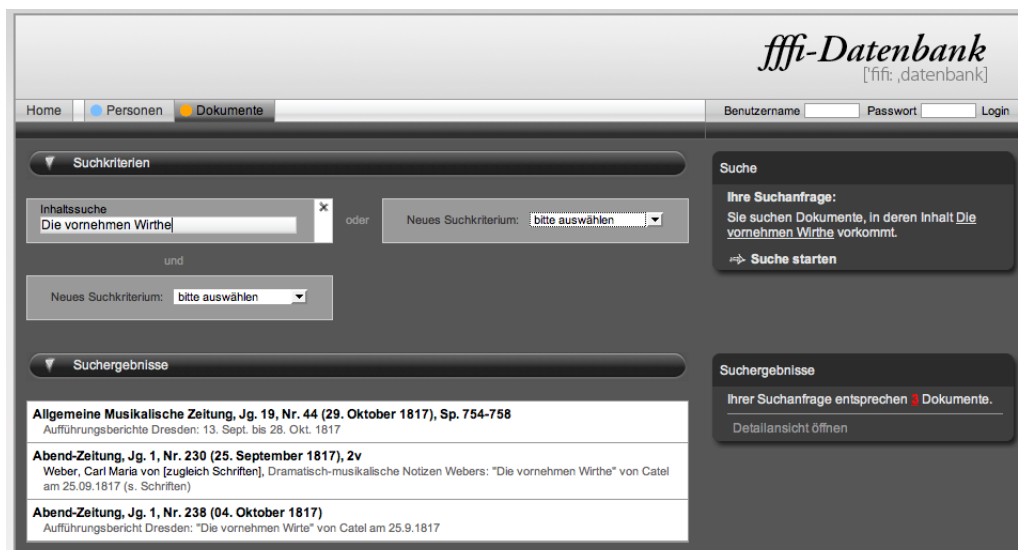


Abbildung 3: Dokumente in der fffi-Datenbank

nun das aktuell in unserer Datenbank befindliche zeigen.⁹ Zunächst fällt (in dem hier nicht abgedruckten Beispiel) der doch deutlich stärker aufgefüllte Bereich der Metaangaben auf, der neben den erklärenden Angaben zum historischen Dokument auch Erläuterungen zum elektronischen Teil, wie z.B. zum Verantwortlichen der Transkription, bereithält. Im Inhaltsbereich des Dokumentes sind Auszeichnungen struktureller Art, wie z.B. Überschriften oder Absätze, auf die Darstellung bezogene Auszeichnungen, wie kursive, unterstrichene oder gesperrte Schrift, und inhaltliche Auszeichnungen, z.B. zu Personen, Werken, Rollen oder Orten, zu finden.

II.

Lassen Sie mich nun zum zweiten Teil meines Vortrags kommen, der auch für die Techniker und technisch versierten unter Ihnen wieder interessant sein dürfte, da ich nun unsere Anwendungsbeispiele für die eben dargestellten Technologien und Konzepte vorstellen möchte. Zunächst werde ich Ihnen die Personen- und Dokumentendatenbank der Fachgruppe Freie Forschungsinstitute zeigen, in welcher wir unsere Auszeichnungsformen nutzen bzw.

⁹Aus Gründen der Übersichtlichkeit verzichte ich hier auf eine Abbildung des vollständigen Dokumentes. Es ist in der Datenbank der Fachgruppe Freie Forschungsinstitute unter http://www.weber-gesamtausgabe.de/ffi-db/docs/ffi-doc_AZ_1_1817_238_2v bzw. direkt als XML-Datei unter http://www.weber-gesamtausgabe.de/ffi-db/docs/ffi-doc_AZ_1_1817_238_2v.xml zu finden.



Abbildung 4: Geöffnetes Dokument mit hervorgehobenen Personen

nutzen werden.¹⁰ Über die Startseite der fffi-Datenbanken gelangt man zu der Datenbank der Dokumente bzw. der historischen Presstexte. An dieser Stelle lassen sich diverse Suchanfragen über verschiedene Kriterien, wie z.B. einer Suche nach dem Autor, Erscheinungsdatum oder einer Inhaltssuche, an die Datenbank richten, die im Falle eines positiven Ergebnisses der Suchanfrage eine Liste aller Dokumente, die der Suche entsprechen, anzeigt (Abbildung 3). Die Anzeige eines Dokumentes innerhalb der Datenbank löst die in XML codierten Auszeichnungen auf und stellt das Dokument mit den entsprechenden Visualisierungen dar. Die Auszeichnungen zum Inhalt werden erst dann sichtbar, sobald Fragen an das Dokument gestellt werden, wie z.B. welche Personen oder Werke im Dokument erwähnt werden. Dieses Verhalten wird in Abbildung 4 verdeutlicht, hier ist das Dokument aus den vorigen Beispielen mit hervorgehobenen Personen im Text abgebildet. Diese Hervorhebe-Mechanismen basieren auf den inhaltlichen Auszeichnungen im XML-Quelltext des Dokumentes. Zusätzlich zu der Möglichkeit, die einzelnen Objekte im Text graphisch herauszustellen, lassen sich auch Funktionalitäten mit diesen Objekten verknüpfen.

Wie Sie aus dem Vortrag von Herrn Veit wissen, sind die Informationen, mit denen er täglich bei der Arbeit rund um Weber zu tun hat, stark verstreut und in vielen Bereichen zu finden; Briefe enthalten Informationen zu Tagebucheinträgen, Einträge in Webers Tagebüchern erklären Vorkommnisse, die wir auch in der zeitgenössischen Presse lesen. So könnte man die inhaltliche Auszeichnung der Dokumente auch z.B. auf diese anderen zeitgenössischen Texte ausdehnen und auch an diesen Stellen Querverweise zu erzeugen. Da die Datenbanken der Fachgruppe noch im Aufbau sind und sich noch in einem sehr frühen Stadium der denkbaren Komplexität und Vollständigkeit befindet, sind die ersten und bislang einzigen Verknüpfungen, die tatsächlich mit Funktionalität versehen sind, die Personen innerhalb der Dokumente. Die Personen sind neben einer Auflösung des Namens zu der im Register üblichen Namensform zum Teil auch mit Identifikationsnummern versehen (`<person fffi="fffi-person_2148" reg="Wilhelmi, Georg Wilhelm">Wilhelmi</person>`), die eine direkte Verknüpfung zu den in der Personendatenbank gespeicherten Personen erlauben, ohne dass etwa von Hand ein Link gesetzt werden müsste. Durch einen Mausklick auf eine Person im Dokument mit Identifikationsnummer gelangt man direkt in die Detailansicht der Person, in welcher nähere Informationen wie z.B. die Lebensdaten oder der Beruf zusammengefasst sind.

¹⁰An dieser Stelle erfolgte im gesprochenen Vortrag eine Live-Präsentation der Datenbanken, welche für den besseren Vergleich gegenüber den eingefügten Abbildungen auch im Internet unter <http://www.weber-gesamtausgabe.de/fffi-db> zu finden ist.

Auch Personen, die in den Dokumenten noch nicht eindeutig identifiziert wurden, sei es mangels eindeutiger Hinweise auf eine bestimmte Person oder aus Mangel an Zeit, können mit Hilfe der "einfachen" Auszeichnung als Person schon einen Mehrwert generieren; so lässt sich mit einem Klick auf diese Personen direkt eine Suche in der Personendatenbank starten. Wenn diese Methoden der Auszeichnung konsequent durchgeführt werden und das Material in Form von verschiedenartigen Objekten stetig zunimmt, entsteht auf diese Weise ein komplexes und stark verzahntes Konstrukt, welches sich durch zusätzliche Verfeinerung der Beziehungen einmal in ein so genanntes Semantic Web überführen ließe.

III.

Ich möchte mit Ihnen noch einen weiteren Aspekt der elektronischen Datenerhaltung diskutieren und zwar die Nutzung von Informationen und Strukturen durch Dritte. Gemeint ist damit, die Möglichkeit zu schaffen, Datenbestände nicht nur für einige wenige zur Verfügung zu stellen, sondern auch für andere Personen und Programme Informationen nutzbar zu machen. So sollte immer die Erstellung eines offenen Standards für die Strukturierung der Inhalte angestrebt werden, solange nicht schon ein Standard existiert, der den eigenen Anforderungen, vielleicht auch mit wenigen Anpassungen, schon genügen würde. Solange diese Prämissen für die eigenen Arbeiten eingehalten werden, können langfristig Kosten eingespart werden, die Strukturen wären verständlicher und unter Umständen einfacher, da nicht nur einige wenige an der Gestaltung mitgewirkt haben und mit ihnen arbeiten. Auch wären die Anforderungen an die Struktur seitens der einzelnen Projekte vielfältiger, sodass auch verschiedenartige Programme entstehen würden, die aber mit den gleichen Daten arbeiten könnten. Innerhalb der Datenbanken der Fachgruppe sollen im Bereich der Dokumente die Auszeichnungsstandards der TEI¹¹ in Kombination mit dem Metadatenstandard Dublin Core¹² genutzt werden. Um die Informationen auch einem breiten Publikum zugänglich zu machen, sind sie auf verschiedenen Wegen im Internet abrufbar. Im Augenblick sollten die Datenbanken über die Internetseite der Carl-Maria-von-Weber Gesamtausgabe¹³ aufgerufen werden, die unter der Rubrik "Personen-DB" einen direkten Link bereit hält. Zusätzlich dazu existieren Schnittstellen, die den direkten Zugriff auf die XML-Inhalte erlauben. Das Schema für diese Schnittstellen setzt sich aus dem bekannten

¹¹Internetauftritt der TEI: <http://www.tei-c.org>

¹²Präsenz der Dublin Core Metadata Initiative: <http://dublincore.org/>

¹³Internetauftritt: <http://www.weber-gesamtausgabe.de>

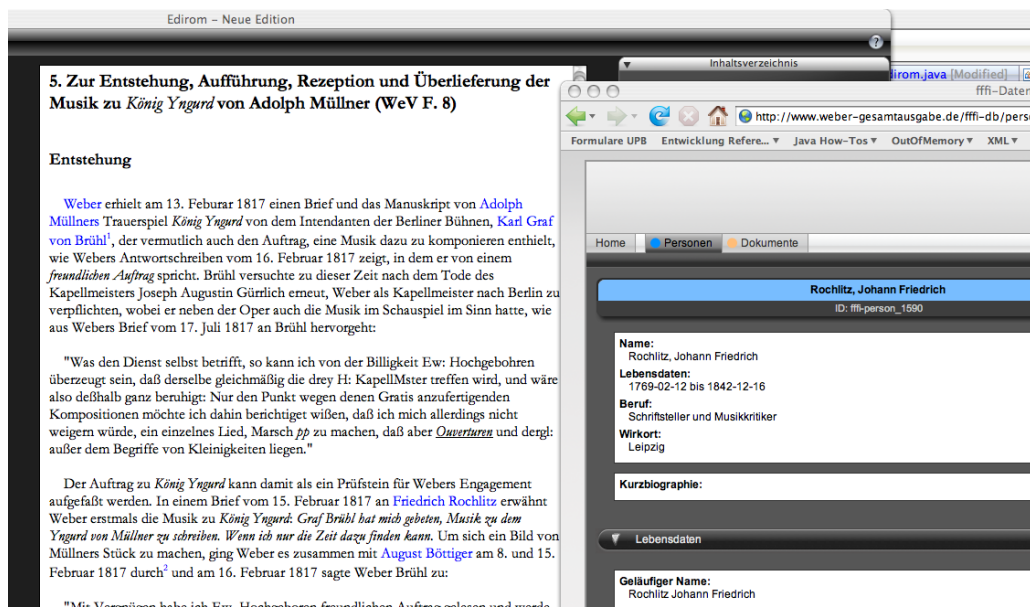


Abbildung 5: Aufruf der Personen-Datenbank aus der *Edirom*

Link zur Datenbank und Angaben zum angeforderten Inhalt zusammen. So ist unter dem Link http://www.weber-gesamtausgabe.de/ffi-db/docs/ffi-doc_AZ_1_1817_238_2v.xml das Dokument (gekennzeichnet durch den Abschnitt "/docs/") mit dem Identifikationskürzel "ffi-doc_AZ_1_1817_238_2v" direkt abrufbar.¹⁴

Eine weitere Möglichkeit, die Informationen auf andere Weise zu nutzen als über die Internetpräsenz, bieten wir mit unserer Editionssoftware *Edirom* an. Zunächst sind lediglich kleinere Versuche zur Vernetzung der *Edirom* mit den Datenbanken durchgeführt worden, um die Möglichkeiten und Anforderungen zu klären; diese Versuche zeigen zunächst eine sehr lose Verknüpfung von im Text enthaltenen Personen und Dokumenten mit den jeweiligen Pendanten in den Datenbanken. Abbildung 5 zeigt diesen Querverweis in Form eines Aufrufs der Personen-Datenbank in einem Browser-Fenster. Vorstellbar ist aber, dass ein Editor bei der Erstellung einer Edition auf die in den Datenbanken zugänglichen Dokumente für die Recherche zugreifen und sie bei ausreichender Relevanz für die Edition sogar in die mit der *Edirom* publizierte Edition einbinden kann. Der Leser der Edition erhält somit Materialien, die in dieser Vollständigkeit und Fülle in einem gedruckten Band nicht vorstellbar wären. Durch die Einbindung eines Dokumentes aus den

¹⁴Für Personen setzt sich die Anfrage in gleicher Weise zusammen, nur dass anstelle des Kennwortes "docs" das Wort "person" eingefügt werden muss. (z.B. http://www.weber-gesamtausgabe.de/ffi-db/person/ffi-person_1819.xml).

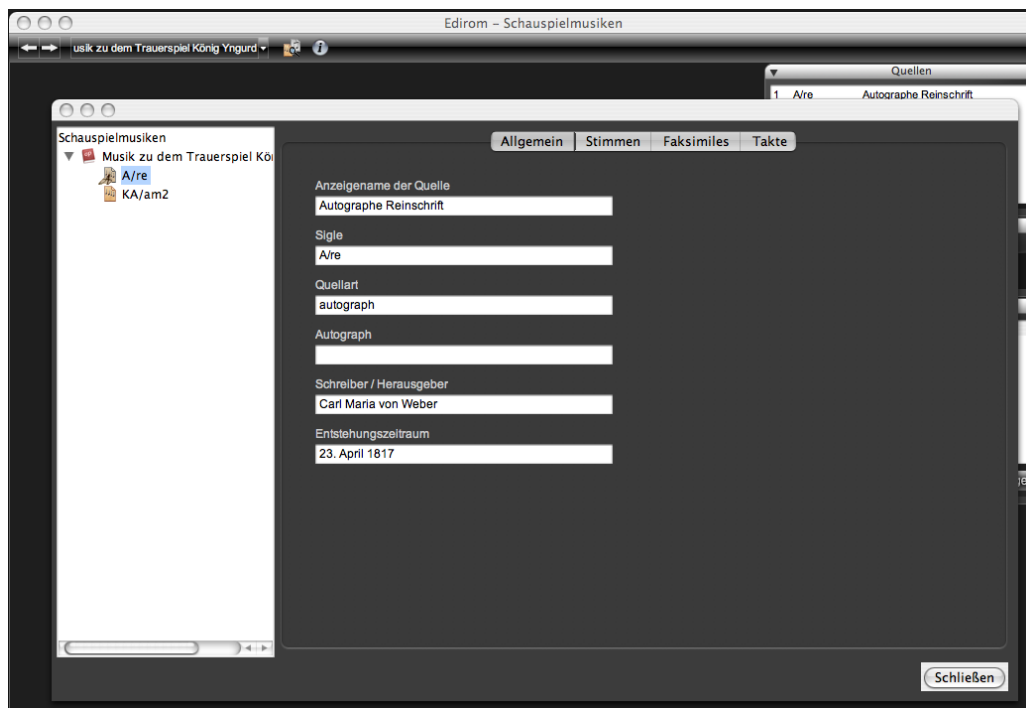


Abbildung 6: Angaben zur Quelle A/re in der *Edirom*

Datenbanken entsteht ein zusätzlicher Mehrwert, der erst wirklich deutlich wird, wenn man daran denkt, dass das Dokument auch in der *Edirom* noch mit dem "Original" in der Datenbank verknüpft bleibt, der Leser also die Möglichkeit hat, über die Auswahl des Editors hinaus eigene Recherchen zu betreiben.

IV.

Ich hoffe, ich konnte Ihnen mit den hier vorgestellten Möglichkeiten und Visionen verdeutlichen, dass es sich durchaus lohnt, diese Technologien einzusetzen, dass also der Mehraufwand für diese Art der Aufbereitung durch den sich ergebenden Nutzen gerechtfertigt ist. Aber "Mehraufwand" ist hier das Schlüsselwort, zu welchem ich im letzten Abschnitt meines Vortrages kommen möchte. Ich habe Ihnen bis jetzt ausschließlich fertige XML-Dateien präsentiert und Ihnen die Arbeit und den Zeitaufwand für die Erstellung derselben verschwiegen. Grundsätzlich ist davon auszugehen, dass die Erzeugung einer XML-Datei bzw. deren Bearbeitung nicht manuell in einem Texteditor oder XML-Editor geschehen sollte, sodass Sie hoffentlich nie mit diesem für Sie unter Umständen kryptisch anmutenden Quelltext einer XML-Datei in Berührung kommen müssen. Im Idealfall entwickeln einige wenige

```

<edition id="24562047">
  <name>Schauspielmusiken</name>...
  <works>
    <work id="1">...
      <sources>
        <source id="1">
          <name>Autographe Reinschrift</name>
          <siglum>A/re</siglum>
          <category>autograph</category>
          <autographParts />
          <writerName>Carl Maria von Weber</writerName>
          <creationDate>23. April 1817</creationDate>
          <sortNo>1</sortNo>
          <voices />
          <facsimiles>...</facsimiles>
        </source>
        <source id="2">...</source>
      </sources>
    </work>
  </works>
</edition>

```

Abbildung 7: Quelltext der Eingaben aus Abbildung 6

Fachkundige eine Struktur für die zu erzeugenden Dateien und bieten ein Programm für die Bearbeitung an. Diese Programme sollten jegliche Möglichkeiten des zugrunde liegenden XML-Formates in eine für Laien intuitiv zu nutzende graphische Oberfläche übersetzen. Die von uns entwickelte *Edirom* ist ein solches Programm, welches zur Generierung – von natürlich speziell auf die Inhalte von Editionen abgestimmte – XML-Dateien genutzt werden kann. In Abbildung 6 sind die Meta-Angaben zu einer Quelle innerhalb einer Edition zu sehen, sie erstrecken sich von dem Anzeigenamen über die Sigle, die Art der Quelle bis hin zum Schreiber und zum Entstehungsdatum. Diese Angaben werden durch Eingabehilfen unterstützt in die abgebildete Oberfläche geschrieben; der dadurch erzeugte XML-Quelltext ist in Abbildung 7 zu finden. Jede Änderung der Edition innerhalb der *Edirom* erzeugt in der zugrunde liegenden Datei XML-Code, der über eine Funktion in eine externe Datei exportiert werden kann. Die strukturierte Repräsentation von eingegebenem Text in einer XML-Datei ist, wie Sie an diesem

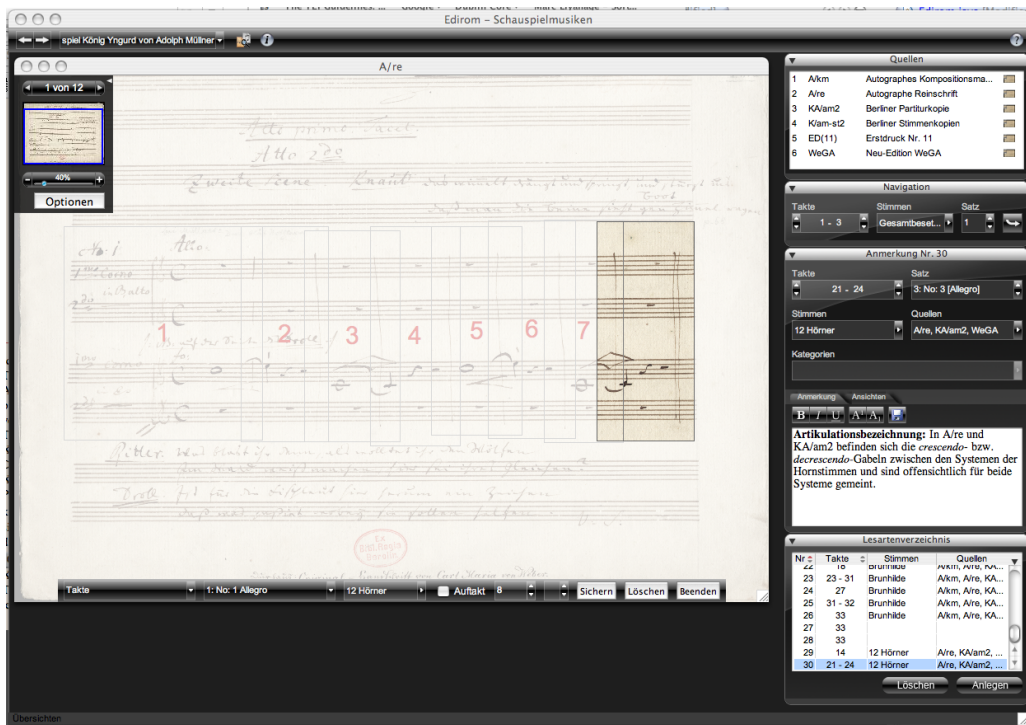


Abbildung 8: Eingabe von Taktinformationen in der *Edirom*

Beispiel gesehen haben, sehr einfach möglich. Schwieriger wird es, wenn neben den textuellen Inhalten auch Bilder und Angaben zu diesen innerhalb einer XML-Datei festgehalten werden müssen. So werden die Takte auf den einzelnen Faksimile-Seiten der Quellen mit einem Rahmen markiert, damit diese Positionsinformationen z.B. für die parallele Kollationierung der Quellen oder für die Bestimmung des Anzeigebereiches für eine Anmerkung des Kritischen Berichts genutzt werden können. Das Aufziehen eines Takt-Rahmens auf einem Faksimile wird dem Editor durch graphische Hilfsmittel ermöglicht, die unabhängig von der Position oder den Zoom-Faktoren des Bildes die korrekten Eckpunkte des Taktes ermitteln können. Diese Oberfläche ist in Abbildung 8 beispielhaft am Takt 8 der autographen Reinschrift der Beispieldition dargestellt. Zu sehen sind die Einstellmöglichkeiten für die Angaben zum Takt einschließlich der Angabe der umschließenden Stimmen für unterschiedlich genaue Eingabestufen. Der durch diese Eingabe erzeugte XML-Text ist in Abbildung 9 enthalten; erkennbar ist der Takt 8, repräsentiert durch das Element "bar" mit der Identifikationsnummer 83. Gespeichert werden neben dem Namen bzw. der Nummer des Taktes auch die Position des Taktes in der sortierten Reihenfolge sämtlicher Takte der Quelle für diesen Satz, eine Referenz auf den Satz selbst, die enthaltenen

```

<source id="7">
...
<facsimiles>
  <facsimile id="1">
    <name>1r</name>
    ...
    <bars>
      <bar id="83">
        <name>8</name>
        <sortNo>8</sortNo>
        <movement>1</movement>
        <coord x="164" y="648" width="602" height="647" />
        <voices>
          <voice id="1" />
        </voices>
      </bar>
    ...
  
```

Abbildung 9: Quelltext der Taktangaben inkl. der Positionierung (Auszug)

Stimmen und natürlich die Koordinaten des Rechtecks auf dem Faksimile. Zukünftig sollen sämtliche Textelemente einer Edition innerhalb der *Edirom* herstellbar sein, sodass keine externen Textverarbeitungswerkzeuge für die Gestaltung der Texte und keine XML-Kenntnisse für die inhaltliche Auszeichnung benötigt werden.¹⁵ Selbstverständlich werden die Texte Querverweise enthalten, aber auch Bilder einbinden können, die nicht zwingend in ihren Abmessungen statisch sein müssen; denkbar wäre eine Integration eines mit Bewegungs- und Zoom-Funktionen ausgestattetes Bildelement. Neben dieser technisch recht einfachen Einbindung der Textteile soll langfristig auch die Erstellung und Visualisierung von Notentext innerhalb der *Edirom* möglich sein. Sie wird die Digitalisierung der historischen Quellen in eine für den Computer nutzbare Form, aber auch die Erstellung des Referenznotentextes umfassen.¹⁶

¹⁵Für die Struktur der XML-Dateien werden soweit als möglich die Vorgaben der TEI genutzt werden.

¹⁶Nähere Erläuterungen zu der digitalen Codierung von Noten finden Sie im Vortrag von Johannes Kepper im Rahmen desselben Kolloquiums.