

Dokumentation mit XML im Germanischen Nationalmuseum Nürnberg

In der deutschen Museumswelt ist die Dokumentation im Gegensatz zum Bibliothekswesen von einer gewissen Vielfalt und Individualität der Erfassungsformate geprägt. Einen allgemein anerkannten Standard – wie RAK innerhalb der Bibliothekswelt – gibt es nicht. Allenfalls MIDAS hat über die Projekte von Foto Marburg innerhalb der Museen eine weitere Verbreitung erfahren. Häufig wird dieser facettenreiche Zustand musealer Dokumentation als Missstand empfunden, und in der Tat wäre es nicht schlecht, wenn man sich innerhalb der deutschen Museen auf ein gemeinsames Format einigen könnte. Grundsätzlich sollte man sich aber hüten, Vielfalt der Erfassungsformate a priori als Missstand aufzufassen. Die vorhandene Vielfalt ist m. E. auch Ausdruck von Unterschieden zwischen Bibliothek und Museum.

Katalogisierung / Objektdokumentation

Auch wenn das allgemeine Ziel des Sammelns, Bewahrens und Erschließens der Museen nicht all zu weit vom Auftrag der Bibliotheken entfernt zu sein scheint, ergeben sich in der Praxis doch zum Teil ganz erhebliche Handlungsunterschiede. Das Museum verwaltet in aller Regel eine Vielzahl von Objektarten, die in unvorhersehbaren Bezugsmustern auftreten können. Noch tragender ist der Unterschied, dass museale Information einen völlig anderen Charakter als in der Bibliothek aufweisen kann. Denn welches Bibliothekssystem wurde je mit der für unverzichtbar gehaltenen Anforderung ausgeschrieben, den Lebenslauf einer Person mit deren Rolle an einem Herstellungsvorgang eines Objektes innerhalb eines zeitlichen und geographischen Bezugs darstellen zu müssen?

Aus Sicht des Nutzers ist es das vorrangige Ziel eines Bibliothekssystems, einen Zugang zum Standort eines Informationsträgers in Form eines inhaltlichen Verweises zu ermöglichen, aber sicherlich nicht, die Information des Informationsträgers selbst darzustellen oder gar inhaltlich auszuzeichnen, wie dies zumindest in wissenschaftlich orientierten musealen Einrichtungen gefordert wird. Das soll natürlich nicht heißen, dass die Standortfrage nicht auch im Museum von Bedeutung ist, aber darüber hinaus spielen Fragen der wissenschaftlichen Nutzung der Information hinsichtlich eines erweiterten Objektbezugs eine zunehmend wichtige Rolle. Museale Information findet daher konsequent weitergedacht eher Anknüpfungspunkte im Umfeld der *Digital Libraries* als im traditionellen Bibliotheksbereich.

Die Vielfalt in den Erfassungsformaten der Museen ist nicht zuletzt auch Ausdruck der unterschiedlichen

Anforderungen der Forschungsdisziplinen, die in den Museen vertreten sind. Der dokumentarische Nutzen wird von Forschungsbereich zu Forschungsbereich ganz unterschiedlich gewichtet. Allein die so unverfänglich erscheinende Abnahme von Messdaten an einem Objekt kann völlig unterschiedliche Bedeutungen haben. Am GNM sind diese disziplinären Unterschiede auf der Ebene der wissenschaftlichen Sammlungen sehr direkt greifbar. Diese Unterschiede innerhalb der Sammlungen erzeugen am *Germanischen Nationalmuseum (GNM)* die praktische Notwendigkeit, neben dem eigenen, auf Verwaltungszwecke optimierten Standard auch andere Beschreibungsstandards zuzulassen. Was natürlich die nicht trivial zu lösende Frage aufwirft, wie dieser Vielfalt entsprochen werden kann, ohne im Chaos zu versinken?

IST-Analyse: Welche Information für welchen Zweck?

Eine Informationsanalyse am GNM zeigt zwei zunächst nur heuristisch trennbare Nutzungshorizonte der Information. Dies ist einerseits Information, die im wesentlichen administrativen Zwecken dient und andererseits Information, die eine wissenschaftliche Erschließung der Objekte zum Ziel hat. Diese Nutzungshorizonte sind allerdings nur scheinbar zu trennen, da sich in Wirklichkeit die Informationsinhalte vielfältig überschneiden, was in der Organisation der Datenpflege über Abteilungsgrenzen hinweg problematisch ist. Dennoch gelten für beide Informationsbereiche unterschiedliche Bedingungen. Administrative Information ist eine gut determinierbare, vorhersehbare Information mit einem vergleichsweise geringen Grad an Komplexität. Im konventionellen Umfeld ergeben sich Hinweise auf diese Art der Information, wenn mit sehr dezidierten Formblättern Einträge erfasst werden, die nur wenig Freiraum in Form von großen Textblöcken bieten. Diese Art von Information lässt sich in aller Regel sehr einfach mit relationalen Datenbanken verwalten. Ein Beispiel wären Ausleihvorgänge von Objekten. Im Gegensatz dazu steht die wissenschaftliche Erschließung der Objekte: Die hier auftretende Information ist schwierig determinierbar, vielfach unvorhersehbar und von einer hohen Komplexität. Formblätter spielen eine untergeordnete Rolle und wenn, bieten sie meist grosse, beliebig beschreibbare Textbereiche an. Information dieser Art wird meist in Form von variablen Textdateien erzeugt.

Die konventionelle Art und Weise am GNM mit Information umzugehen, weist, was Variabilität und Nutzung von Information angeht, verschiedene Parallelen zum Internet auf, was nahe legt, zumindest eine

Siegfried Krause
(Germanisches
Nationalmuseum,
Nürnberg)

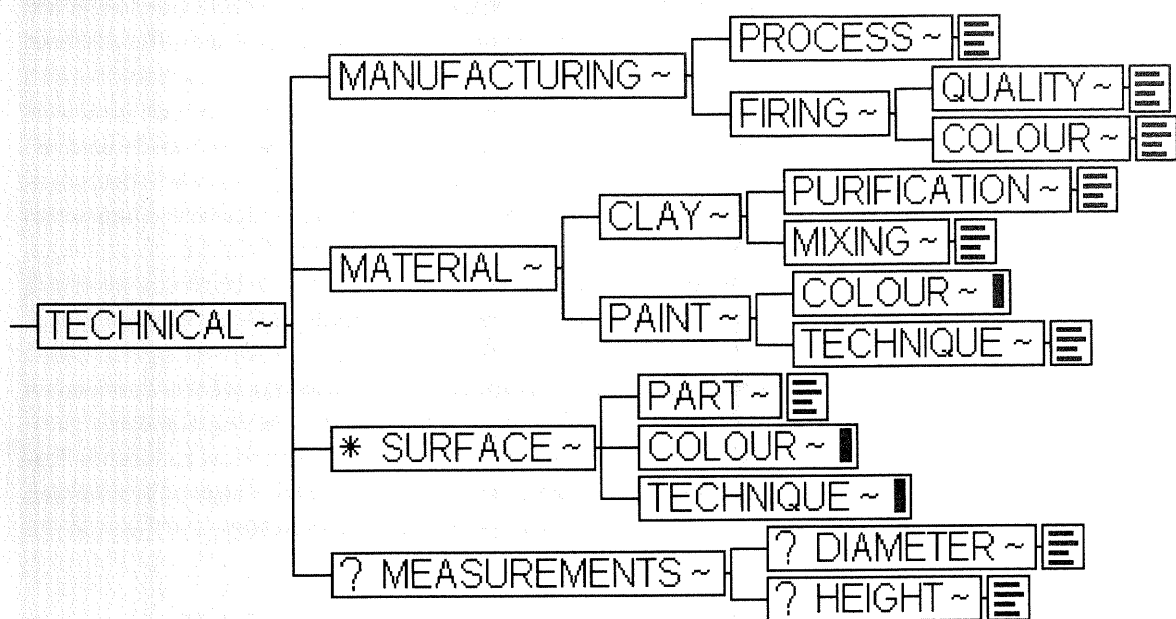


Abbildung: Graphische Darstellung der DTD

technische Problemlösung auch dort zu suchen. So begreift der Mitte letzten Jahres begonnene Realisierungsansatz zum Objektdokumentationssystem des Germanischen Nationalmuseums das Museum selbst als Teil des INTERNET. Alle zum Einsatz kommenden Programme und Datenstrukturen werden auf ihre Integrierbarkeit abgestimmt. In der Konsequenz wird die beschriebene Divergenz der inhaltlichen Nutzungshorizonte der Information technisch in einer zweigleisig dezentralen Datenhaltung umgesetzt.

Ab Mai diesen Jahres wird die zweite Phase des Projekts beginnen, deren Ziel es ist, die verschiedenen Informationsarten zu integrieren. Administrative Information wird in einer relationalen Datenbank abgelegt, während die erschließende Information, ähnlich wie bisher, in Dokumentform erstellt und verwaltet wird. Der Vorteil ist, dass jetzt jede Technologie ihre Stärken kompromisslos ausspielen kann. Strukturierbare Information, wie sie im Zugangsregister enthalten ist, oder Ausleihdaten der Objekte werden in einem DBMS (Datenbank Management System) verwaltet. Aber bereits Inventardaten werden in Dokumentform erfasst und in einer speziellen XML-fähigen Textdatenbank abgelegt. Die Ergebnisse der Recherche werden in Form virtueller Dokumente entsprechend der Recherche und den Berechtigungen des Mitarbeiters über einen INTRANET-WEB-Server dargestellt. In diesem Zusammenhang drückt der Begriff *virtuelles Dokument* nur aus, dass das am Bild-

schirm dargestellte Dokument aus unterschiedlichen Informationsquellen gespeist wird: Das Dokument existiert nicht unbedingt in der angezeigten Form, sondern wird immer für den entsprechenden Bedarf generiert. Auf die Datenbanktechnologie soll im Folgenden nicht näher eingegangen werden. Bei weitem interessanter sind die Vorgänge um die Dokumente, die in Form einer Dokumentenverwaltung organisiert werden, und damit ist auch klar, dass es um strukturierte Dokumente geht.

Strukturierung mit XML

In den letzten zwei Jahren wurde mit XML (Extensible Markup Language) durch das W3C (World Wide Web Consortium) und auch größere Firmen eine Technologie in den Vordergrund gerückt, die in einem engen Zusammenhang mit HTML (HyperText Markup Language) steht. Auch auf die Gefahr hin, dass Vergleiche hinken, lässt sich die Beziehung von HTML zu XML vielleicht am einfachsten in dem Verhältnis eines Programms zu einer Programmiersprache fassen. Mit einer Programmiersprache können beliebig viele verschiedene Programme für unterschiedlichste Zwecke geschaffen werden. Aber das fertige Programm erledigt nur die Aufgabe, für die es geschaffen wurde. Und mit XML verhält sich dies entsprechend. XML ist eine sog. Metasprache, mit der sich der Zweck eines Typs von Dokument definieren lässt. HTML ist z.B. ein solcher Dokumenttyp. HTML hat allein den Zweck, Do-

kumente im WWW darstellbar zu machen. Spezifischer formuliert: XML ist eine Sprache mit der sich die Syntax eines Dokumenttyps formalisieren lässt. Die Formalisierung der Syntax eines Dokumenttyps geschieht durch die Definition einer sog. Document Type Definition (DTD). Eine DTD ist eine Sammlung von abstrakten Koordinationsregeln. Beispielsweise kann eine solche DTD die Regel beinhalten, dass jedes Dokument, das eine Instanz der Klasse Inventardokument präsentiert, ein und nur ein Element zur Auszeichnung der Inventarnummer enthält und dass jedes Dokument mit diesem Element beginnen muss. Ein Dokument mit zwei Inventarnummern würde nicht zum definierten Dokument-Typ der Inventardokumente gehören und stellt damit auch kein Inventardokument dar.

Die Nutzung von DTDs ist nicht nur auf die Verwendung in eigenen Dokumenten oder auf den eigenen Institutionsrahmen beschränkt. Diese formalen Beschreibungen können auch einem via Internet weltweit verbreiteten internationalen Standard entsprechen, der vom XML-Editor direkt aus dem Internet geladen werden kann. Mittlerweile werden auf diese Weise bereits verschiedene hoch spezialisierte DTDs angeboten. Dazu zählen natürlich HTML aber auch weniger bekannte DTDs wie SMIL (Synchronized Multimedia) im Multimediabereich oder MathML, eine mathematische Beschreibungssprache, und natürlich die im Bibliotheksbereich zum Einsatz kommende RDF-Spezifikation (RDF = Resource Description Framework) in Verbindung mit dem Dublin Core Model.

Die technische Umsetzung der Informationserfassung mit XML-basierten Werkzeugen gestaltet sich im Verhältnis zur notwendigen inhaltlichen Diskussion vergleichsweise einfach. Eine entsprechende DTD wird zunächst von einem (validierenden) XML-Editor

eingelassen. Damit stehen im Editor nur die in der DTD definierten Elemente zur Auszeichnung des Inhalts eines Dokuments zur Verfügung. Der ganze Vorgang ähnelt sehr stark der Arbeit mit einer Textverarbeitung, nur dass anstatt eines layout-bezogenen Markups ein strikt inhaltsbezogenes Markup erzeugt wird. XML-Dokumente kann man einfach als *inhaltlich formatierte* Dokumente verstehen. Durch die strikte Festlegung auf eine inhaltliche Auszeichnung (Markup) des Dokuments wird über die eigentliche Nutzungsform des Inhalts zum Teil erst durch die jeweils gewählte Darstellungsform des Dokuments entschieden. Das fertige Dokument wird abschließend gegen die DTD geprüft (validierend geparkt). Soweit die Beschreibung des Vorgangs. Ein einfaches Beispiel mag dies besser verdeutlichen. Die in der Abbildung dargestellte Struktur ist ein Ausschnitt aus einer möglichen DTD zur Beschreibung von Keramik. Die Struktur verdeutlicht verschiedene Eigenschaften von XML-Elementen (sog. Tags). XML-DTDs sind vor allem hierarchisch aufgebaut. Dies macht Begriffsunterscheidungen einfach und ermöglicht qualifizierte Recherchen. Obwohl in diesem Beispiel der Begriff *Colour* mindestens zweimal auftaucht, wird die jeweilige Bedeutung einmal durch die hierarchische Ebene *Paint* (Malfarbe) und das andere Mal durch die hierarchische Ebene *Surface* (Gefäßoberfläche) eindeutig festgelegt. Außerdem können XML-Tags, wie auch in HTML üblich, durch Eigenschaften innerhalb der Element-Tags zusätzliche Informationen transportieren, was allerdings hinsichtlich des Informationsdesigns nicht ganz unproblematisch ist. Durch die DTD wird die Kardinalität der Elemente festgelegt. Beispielsweise wäre das Element *Surface* in diesem Dokumenttyp ein beliebig oft wiederholbares Element. Das Element *Measurements* ist dagegen ein optional vorkommendes Element und das Element *Manufacturing* muss im Dokument ein einziges Mal vorkommen.

Innerhalb eines XML-Dokuments könnte die zugehörige Dokument-Instanz etwa so aussehen:

```
<technical>
  <manufacturing>
    <process>Töpferscheibe</process>
    <firing>
      <quality>hart</quality>
      <colour>grau</colour>
    </firing>
  </manufacturing>
  <material>
    <clay>
      <purification>geschlämmt</purification>
      <mixing>Stroh</mixing>
    </clay>
  </material>
</technical>
```

```

<paint>
  <colour>rot</colour>
  <technique>Kreide</technique>
</paint>
</material>
</technical>

```

In diesem Beispiel fehlen Angaben zur Oberfläche des Gefäßes, was zulässig ist, da die Kardinalitätsregel nur besagt, dass dieses Element keinmal, einmal oder mehrmals auftreten darf. Angaben zum Material sind dagegen obligat.

In allen neuen Versionen der bekannten Internetbrowser (Netscape, Internet Explorer) können XML-Dokumente *in natura* oder über die Verbindung mit Style Sheets und Scripts dargestellt werden. Ohne Style Sheet entspricht die Ansicht dem gezeigten Inhaltsschema. Style Sheets bilden eine Möglichkeit, XML-Dokumente mit einem spezifischen Darstellungsformat zu versehen.

Bisher ausgespart wurde der Bereich der Plausibilitätskontrolle, die mit XML-Syntaxregeln nicht abgedeckt wird. Allerdings gibt es relativ einfache technische Lösungen für dieses Problem. Am GNM müssen zwei Wege parallel beschritten werden, da XML-Dokumente nicht nur online sondern auch offline erstellt werden können.

Ist der Rechner online, kann der am GNM verwendete XML-Editor ein kleines Fremdprogramm (ActiveX-Library sog. In-Process Server) einladen, das eine Verbindung zur Dokumentdatenbank herstellt und gegebenenfalls eine neue Inventarnummer vom System generiert vorschlägt oder aber die von Hand vergebene Inventarnummer wird auf Korrektheit geprüft.

Sowohl für ein offline als auch ein online generiertes Dokument gilt, dass während des Eincheckens in die Dokumentdatenbank die XML-Syntax des Dokuments überprüft wird (sog. Validierender Parser) und in einem anschließenden Schritt die sog. Plausibilitäten kontrolliert werden. Vom System wird im gleichen Zuge ein Protokoll generiert und dem *publizierenden* Mitarbeiter wie auch dem Systemadministrator zugeschickt. Dies ist z.B. hinsichtlich der Wortschatzkontrolle von Interesse, denn ein neu eingeführter Term ist ja nicht automatisch ein Fehler, sondern gilt zunächst als ein Vorschlag, den Thesaurus entsprechend zu erweitern, was in der Folge geklärt werden muss.

Die Gründe, diesen vorgeführten Weg einzuschlagen, liegen vor allem in der Flexibilität der Lösung begründet. Zum einen steht nicht zu erwarten, dass eine DTD in unveränderter Form die Zeit übersteht. Sie wird ergänzt und geändert. In reinen Datenbanksystemen ist damit häufig ein erheblicher Reorganisationsaufwand verbunden, nicht so in den Dokumenten. Auch ist es kein Problem, die historische Veränderung der Information in Dokumenten nachvollziehbar zu dokumentieren, da sie wie ein Blatt Papier funktionieren. Vom Text muss nichts entfernt werden, sondern es wird einfach eine neue Bemerkung angefügt. Auch wird es neben reinen Inventardokumenten andere aufeinander bezugnehmende Dokumentarten geben wie Restaurierungsprotokolle, Archivdokumente, Texte von Bestandskatalogen etc. Zusätzlich besteht am GNM auch die Notwendigkeit, MIDAS-Dokumente zu importieren, was in jedem Fall machbar ist, aber Thema eines eigenen Projekts sein wird. Es ist allerdings weniger ein technisches als ein inhaltliches Projekt.

Das XML-Projekt am GNM

Die Umsetzung des Projekts am GNM hat mit der Entwicklung und Modellierung der Dokumentstruktur begonnen. Seit Herbst 1999 werden XML-Editoren an den Arbeitsplätzen eingeführt. Die zweite, ab Mai 2000 beginnende Projektphase hat die Einführung der Dokumentdatenbank zum Ziel. Erste Recherchemöglichkeiten werden ab Dezember 2000 implementiert sein. Hinzu kommen bis Mitte 2001 eine thesaurus- und verwaltungsunterstützende Funktionalität.

XML-basierte Dokumente verknüpfen den Vorteil strukturierter Information (Syntax eines Dokuments) mit gleichzeitig hoher Flexibilität. Der Vorteil für die Mitarbeiter innerhalb der Sammlungsreferate ist, dass sie in ihrer gewohnten Textumgebung arbeiten können, denn dies ist, wie eingangs bemerkt, einer der Unterschiede zwischen bibliothekarischer und musealer Information. Hervorzuheben ist auch, dass die Semantik der Texte erhalten bleibt, was den wissenschaftlichen Nutzwert deutlich erhöht. Durch die Darstellung der Information in Textform wird zudem

verdeutlicht, dass die Information, wie in geisteswissenschaftlichen Zusammenhängen üblich, als eine Meinung zu einem Sachverhalt aufgefasst werden muss.

Die Einführung der dokumentenbasierten Informationsverarbeitung am GNM drückt die Erwartung aus, dass die modernen Informations- und Kommunikationsmedien auch im Bereich der Kulturwissenschaften zu neuen Arbeitsformen und Methoden führen werden. Der über Jahrzehnte hinweg primäre Verwendungszweck der Dokumentation wurde am GNM wie anderen Orts als ein zutiefst innerer Nutzen begriffen. Der Nutzen war wesentlich bestimmt durch die Sorgfaltspflicht; die Dinge waren verwaltbar.

Diese Situation ist in Veränderung geraten. Das Postulat *Wissen ist Macht* wird in Zukunft ergänzt, wenn nicht ersetzt werden müssen, durch das Postulat *Wissen ist Existenz*. Wissenschaftliche Museen und Einrichtungen werden nicht umhin kommen, die von ihnen erzeugte Information in ähnlicher Selbstverständlichkeit wie Bibliotheken einem breiten und zugleich auch sehr heterogenen Nutzerkreis über moderne Kommunikationswege anzubieten.

Siegfried Krause
(Germanisches Nationalmuseum, Nürnberg)

Die folgenden WWW-Informationsangebote sind nur als Einstiegsinformation gedacht.

– *Allgemeine Informationsseiten zu Markup Sprachen*

<http://www.uni-koblenz.de/~www/sprachen/beschreibung.html>

<http://www.xml.com/pub>

<http://www.w3.org/XML/>

<http://www.mintert.com/xml/trans/REC-xml-19980210-de.html>

<http://www.heise.de/ix/raven/Web/xml/>

– *Spezifische Standards in XML abgelegt*

Resource Description Framework:

<http://www.w3.org/RDF/>

Dublin Core: <http://purl.org/DC/index.htm>

Synchronized Multimedia:

<http://www.w3.org/AudioVideo/>

Mathematical Markup Language:

<http://www.w3.org/Math/>

– *Speziell zu Museen*

CIDOC Conceptual Reference Model (CRM):

<http://www.ics.forth.gr/proj/isst/Activities/CIS/cidoc/>

AMICO Projekt: <http://www.amico.org/home.html>

ERASMUS

Ihr Lieferant für alle Kunstbücher

Bitte richten Sie Ihre Bestellungen und Anfragen an:

ERASMUS BV
P.O. BOX 19140
1000 GC AMSTERDAM
The Netherlands
Tel.: +31-20-627 69 52
Fax: +31-20-620 67 99
E-mail: erasmus@erasmusbooks.nl
Internet: <http://www.erasmusbooks.nl>

Für französische Bücher:
Librairie Erasmus
28, rue Basfroi
75011 Paris, France
Tel.: +33-1-43 48 03 20
Fax: +33-1-43 48 14 24
E-mail: erasmus@erasmus.fr



Erasmus

- Professional
- Traditional
- Academic