

# Extensible Markup Language (XML)

Dirk Fox

*Gateway (Tor, Durchgang) ist ein Vermittlungscomputer, der zwei unterschiedliche, aber gleichartige Kommunikationssysteme verbindet. Dazu setzt er die Kommunikationsregeln (Protokolle) des einen in die des anderen um. So ermöglicht er den Teilnehmern beider Systeme, sich zu verständigen. In unserem „Gateway“ werden Juristen technische und Technikern juristische Begriffe erläutert.*

Die „Extensible Markup Language“ XML ist eine vom W3W-Konsortium spezifizierte Meta-Beschreibungssprache für strukturierte Daten. Darunter werden Dokumente verstanden, die einen hierarchischen und systematisch gegliederten Aufbau besitzen, wie Tabellen oder Karteikarten, Datenbank-einträge oder Formulare.

XML erlaubt eine sehr freie, insbesondere rein syntaktische, d. h. an keine Bedeutung oder Interpretation gebundene Darstellung solcher Datensätze: Jedes Element eines XML-Dokuments wird durch selbst definierbare „Tags“ gekennzeichnet, die den Anfang und das Ende eines Strukturelements, beispielsweise eines Feldeintrags oder einer Tabellenspalte, definieren.

So ließe sich beispielsweise ein Beitrag in DuD Gateway wie folgt strukturiert darstellen:

```
<DuD>
  ...
  <Gateway>
    <Titel>Extensible
    Markup Language (XML)</Titel>
    <Autor>Dirk
    Fox</Autor>
    <Definition>Gateway
    (Tor, Durchgang) ist ein Vermittlungs-
    computer, der ...
    </Definition>
    </Text>...</Text>
  </Gateway>
</DuD>
```

Da die gewählten Tags keine standardisierte Bedeutung haben, hätte in obigem Beispiel statt <DuD> auch <xyz>, statt <Gateway> das Tag <stk4d8z3> stehen können. Eine Bedeutung erhalten Tags erst durch die Interpretation des Programms, dass das strukturierte XML-Dokument auswertet. Die Wahl „menseninterpretierbarer“ Tags ermöglicht auch eine manuelle Bearbeitung von XML-Dokumenten in einem Texteditor; dies kann in vielen Anwendungen (z. B. bei Konfigurationsdateien) von Vorteil sein.

Darstellungen in XML müssen im Wesentlichen den beiden folgenden syntaktischen Anforderungen genügen:

- ♦ Sie müssen immer *wohldefiniert* sein, d. h. die verwendeten Tags müssen je-

weils „geöffnet“ und „geschlossen“ werden, und

- ♦ Tags dürfen nur *streng geschachtelt* werden, sich aber nie „überkreuzen“, d. h. das zuletzt „geöffnete“ Tag muss als erstes „geschlossen“ werden.

XML wird eine große Zukunft bei der Verarbeitung strukturierter Daten, wie z. B. Datenbankeinträgen, und bei Finanz-Transaktionen wie Überweisungen oder EDI-Nachrichten vorausgesagt.

Auch bei Datenschutz und Datensicherheit hält XML Einzug. So setzt das vom W3C-Konsortium standardisierte Webseiten-Markierungsverfahren PICS<sup>1</sup>, das eine wichtige Rolle bei Ansätzen zur Kategorisierung von Webseiten spielt, auf XML [Neun\_99, W3C\_98]. Und auch P3P<sup>2</sup> verwendet XML als Beschreibungssprache für die Darstellung von Privacy Policies [CGMS\_00, W3C\_99].

Mit dem Einsatz von XML zur Darstellung wichtiger öffentlicher Informationen (Labels, Policies etc.) entstand der Bedarf an einem Mechanismus zum Schutz der Integrität der dargestellten Informationen. Vor allem zwei Initiativen zur Integration von digitalen Signaturen in XML-Dokumente sind wichtig:

- ♦ Die Spezifikation von *digital signatures for XML* durch die IETF<sup>3</sup> sowie
- ♦ die Spezifikation von *PICS signed labels* durch das W3C-Konsortium.

Die IETF-Spezifikation erlaubt dabei sowohl die Einbettung der signierten Daten in das die Signatur enthaltende XML-Dokument als auch die Aufnahme einer Referenz (einer URL<sup>4</sup>) auf das signierte Dokument. Die Signatur kann sich dabei auf das gesamte referenzierte Dokument

oder aber auch nur auf gekennzeichnete Teile beziehen.

Für viele mögliche Anwendungen, z. B. Workflow-Systeme, ist die in der Spezifikation vorgesehene Option, Mehrfachsignaturen zu demselben Dokument unterzubringen, besonders wichtig.<sup>5</sup>

Dass mit der Nutzung von signierten XML-Dokumenten im Detail eine Reihe von Schwierigkeiten verbunden sind, wurde in dieser Zeitschrift bereits diskutiert [Schm\_00].

## Literatur

- [CGMS\_00] Ann Cavoukian, Michael Gurski, Deirdre Mulligan, Ari Schwartz: *P3P und Datenschutz*. DuD 8/2000, S. 475-478.
- [GrWä\_99] Grimm, Rüdiger; Wäsch, Jürgen: *XML und IT-Sicherheit*. In: R. Baumgart, K. Rannenber, D. Wähler, G. Weck (Hrsg.): *Verlässliche Informationssysteme*, Vieweg 1999, S. 141-162.
- [Neun\_99] Neundorf, Dörte: *Filtertechnologien zur Reduktion der Jugendgefährdung im Internet*. In: R. Baumgart, K. Rannenber, D. Wähler, G. Weck (Hrsg.): *Verlässliche Informationssysteme*, Vieweg 1999, S. 357-366.
- [Schm\_00] Schmidt, Andreas: *Signiertes XML und das Präsentationsproblem*. DuD 3/2000, S. 153-158.
- [W3C\_98] W3C Initiative: *Platform for Internet Content Selection (PICS)*, <http://www.w3c.org/PICS/>
- [W3C\_99] W3C Initiative: *Platform for Privacy Preferences (P3P)*, <http://www.w3c.org/P3P/>

<sup>1</sup> Platform for Internet Content Selection – PICS.

<sup>2</sup> Platform for Privacy Preferences – P3P, siehe Cranor, Gateway, DuD 8/2000, S. 479.

<sup>3</sup> Internet Engineering Task Force, siehe Fox, *Internet-Standard*, Gateway, DuD 7/2000, S. 417.

<sup>4</sup> Unified Resource Locator; eine „WWW-Adresse“, die angibt, wo ein bestimmtes Dokument (z.B. eine HTML-Seite) zu finden ist.

<sup>5</sup> Zu einer weiteren möglichen Anwendung von signierten XML-Dokumenten in der IT-Sicherheit siehe Geuer-Pollmann/Schweitzer, *Vergleichbarkeit von Policies mittels XML*, in diesem Heft.