



Is There a Role for XML in Health Informatics?

*Frank Tompa
March 18, 2004*

XML = Extensible Markup Language

- ▶ A mechanism for defining and representing information as structured documents for applications on the Internet

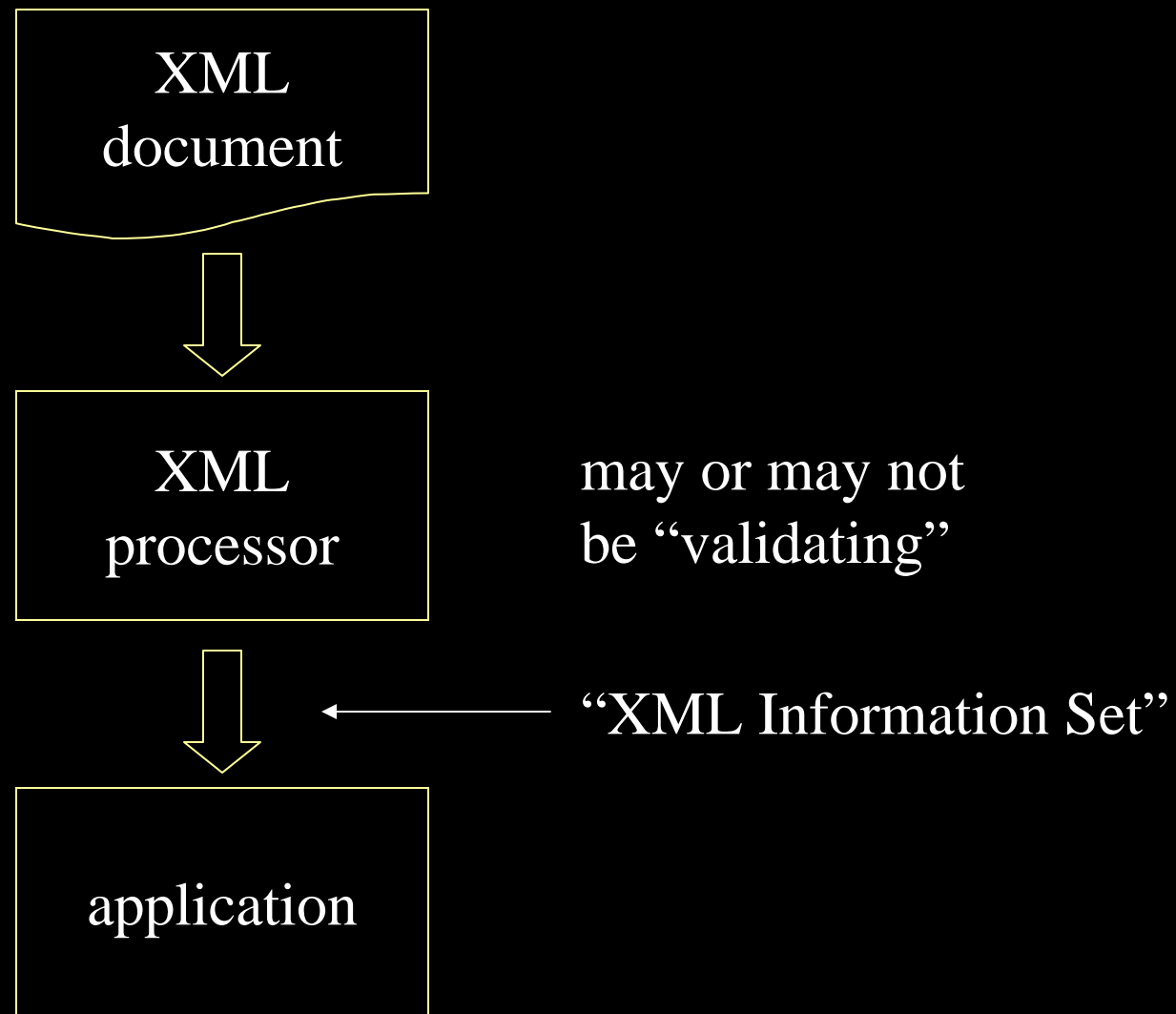
- ▶ Example: Healthcare Informatics

Subcommittee E31.25, 'XML Document Type Definitions (DTDs) for Health Care' is "developing DTDs for health care documents. In October of 1999 we held a joint meeting with the Massachusetts Medical Society and developed draft DTDs for prescriptions, discharge summaries, encounter notes, operative reports, diagnostic imaging reports and the HCFA 1500 form for Medicare reimbursements. We are actively seeking submissions to further the DTD development." [American Society for Testing and Materials, 17 Feb 2000]

XML is a metalanguage, not a specific language.

- ▶ Defines the rules how to mark up a document — does not define the names used in markup
- ▶ Includes capability to prescribe a *Document Type Definition (DTD)* to constrain the markup permitted in a class of documents
- ▶ Includes capability to prescribe a *Schema* to constrain the types of data permitted in various components within a class of documents

Interpreting Tagged Text



Data storage and interchange

- ▶ Encoding of data
 - as a stored data format
 - as an exchange format
- ▶ Encoding of communications protocols
 - designating functions
 - providing parameter values
 - wrapping returned data

- ▶ ISIS European XML/EDI Healthcare Pilot Project (XMLEPR)
- ▶ Open Healthcare Group 'XChart'
- ▶ DocScope: Open Source XML Healthcare Project
- ▶ Health Level Seven XML Patient Record Architecture
- ▶ ASTM XML Document Type Definitions (DTDs) for Health Care
- ▶ The CISTERN Project – Standard XML Templates for Healthcare

See <http://xml.coverpages.org/xmlApplications.html>

- ▶ Health and medical literature
- ▶ Maintenance of patient records
- ▶ Data interchange between independent systems
- ▶ Descriptions of health care procedures and protocols
 - Guidelines
 - Individual instances
- ▶ Health-related information requests

What are my own interests?

- ▶ Data representation, storage, and maintenance
- ▶ Search and display
 - Exact match
 - Information retrieval
- ▶ Data classification / mining
- ▶ Data transformation
 - For data integration or sharing
 - For defining user views of data

What are my own interests?

- ▶ **Specifics**
 - [OED project](#)
 - Text/Relational Database project
- ▶ **Recent and current PhD students' interests**
 - XQuery processing in a Text/Relational environment
 - Specification of structured text transformations
 - Mining topic flows from unstructured text
 - Query optimization in the presence of user views
 - Fine grained, user-managed access control
 - Data integration

What am I looking for?

“I have no data yet. It is a capital mistake to theorize before one has data.”

A. Conan Doyle,

Adventures of Sherlock Holmes,

“Adventure 1: A Scandal in Bohemia,”

p. 3.

What am I looking for?

- ▶ Mix of structured, unstructured, and semi-structured data
 - Not merely plain text
 - Not suitable for relational databases
- ▶ Fairly well-understood application needs
- ▶ Possibly, opportunities for data integration



Thank
you