

Lecture Outline

Metadata

Michael J. Watts

<http://mike.watts.net.nz>

- What is metadata?
- The need for metadata
- XML
- Metadata and XML
- Problems with XML

What is metadata?

- Data about data
- Documentation about data
- Describes a set of data
 - Source
 - Dimensions
 - Variables

What is metadata?

- Describes an object
 - Owner
 - Location
- Metadata has structure
- Metadata can describe structure
 - Database schema

The Need for metadata

- Sufficiency
 - Can an object describe itself?
 - e.g., images
- Scalability
 - Allows for rapid searching
 - Searching metadata fields vs large data files
- Interoperability
 - Can exchange data using mutually agreed metadata formats

XML

- Extensible Markup Language
 - Uses tags to describe data elements
- A standard defined by the W3C
 - Standard for data exchange over networks
- Few predefined elements
 - Minimalist
- A tree structure
 - Parent nodes, subnodes

XML

- Example XML document

```
<address>
<name>
<title>Mr.</title>
<firstname>John</firstname>
<lastname>Smith</lastname>
</name>
<street-address>1000 Main Street</street-address>
<city>Los Alamitos</city>
<state>CA</state>
<zip>90720</zip>
</address>
```

- Steven J. Vaughan-Nichols, "XML Raises Concerns as it Gains Prominence", IEEE Computer, May 2003, pg14-16

XML

- Elements in XML described by a DTD
 - Document Type Definition
- Describes elements
 - Name
 - Type
 - Format
 - Order

XML

- XML document is **Well-formed** if
 - Element tags are matched
 - The tags are closed correctly
- XML document is **Valid** if
 - Structure conforms to the DTD

Metadata and XML

- The elements in an XML document describe the data
 - Elements are metadata
- Grammar in the DTD describe the elements
 - Metadata for the elements
- The flexibility of XML illustrates the use of metadata

Problems with XML

- Large files
 - Tags add bulk
 - Addressed with compression
- Security
 - Files are plain text
 - Addressed with encryption

Summary

- Metadata is data about data
- Needed to help with
 - describing data sets
 - searching for data
 - communicating data
- XML is a metadata format
- XML widely used to transfer data between systems