XPATH
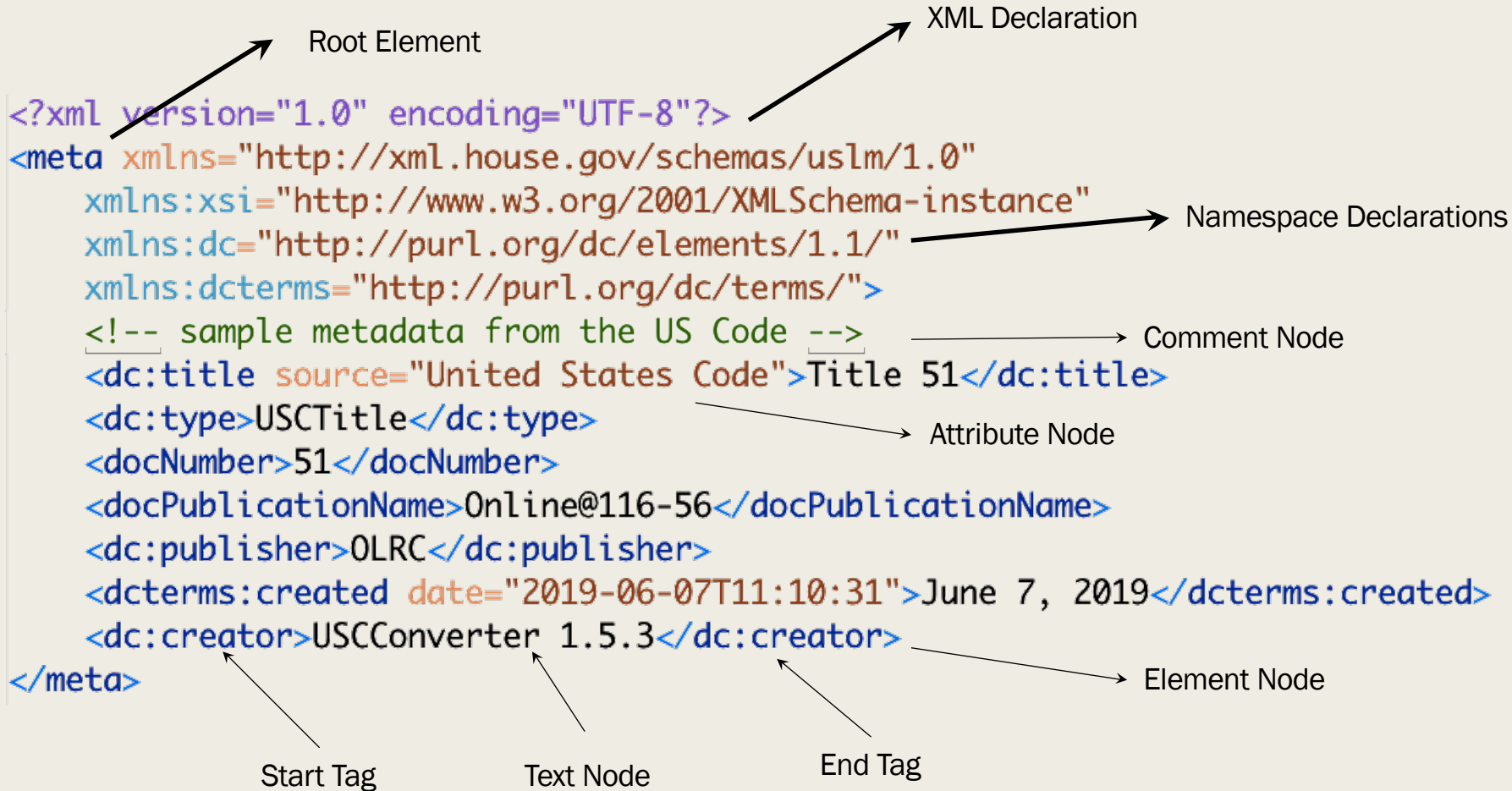
Clifford B. Anderson

Associate University Librarian for Research and Digital Initiatives

"The primary purpose of XPath is to address the nodes of XML 1.0 or XML 1.1 trees. XPath gets its name from its use of a path notation for navigating through the hierarchical structure of an XML document."

— XML Path Language (XPath) 3.0

Root Element → XML Declaration →

```xml
<?xml version="1.0" encoding="UTF-8"?>
<meta xmlns="http://xml.house.gov/schemas/uslm/1.0"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:dcterms="http://purl.org/dc/terms/">
    <!-- sample metadata from the US Code -->
    <dc:title source="United States Code">Title 51</dc:title>
    <dc:type>USCTitle</dc:type>
    <docNumber>51</docNumber>
    <docPublicationName>Online@116-56</docPublicationName>
    <dc:publisher>OLRC</dc:publisher>
    <dcterms:created date="2019-06-07T11:10:31">June 7, 2019</dcterms:created>
    <dc:creator>USCConverter 1.5.3</dc:creator>
</meta>
```

XML Declaration

Namespace Declarations →

Comment Node

Attribute Node

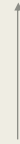Element Node

Start Tag

Text Node

End Tag

Path Expressions

- axis : "defines the 'direction of movement' for the step"
- node test: "selects nodes based on their kind, name, and/or type annotation"
- optionally: a predicate, which filters results according to some Boolean test

/descendant::usc:uscDoc/descendant::usc:meta[dc:title/text() = "Title 17"]

axis

node test

predicate

Forward Axes

- child::
- descendant::
- attribute::
- self::
- descendant-or-self::
- following-sibling::
- following::
- namespace::

/descendant::usc:subsection/child::usc:content

Reverse Axes

- parent::
- ancestor::
- preceding-sibling::
- preceding::
- ancestor-or-self::

/descendant::dc:title/parent::element(usc:meta)

Node Tests

- node() matches any node.

- text() matches any text node.

- comment() matches any comment node.

- namespace-node() matches any namespace node.

- element() matches any element node.

/descendant::usc:title/child::element()/child::usc:p

Abbreviated Syntax

- .   context node

- // descendant-or-self::node()

- @ attribute::

- .. ancestor::node()

- /   child::node()

//
//usc:note[fn:contains(., "enacted")]/following-sibling::usc:date

Predicates

A predicate filters out results from path expressions based on some Boolean test.

//usc:meta[dc:title/text() = "Title 17"]

//usc:section[usc:heading[. = "Definitions"]]/usc:paragraph/usc:content/node()