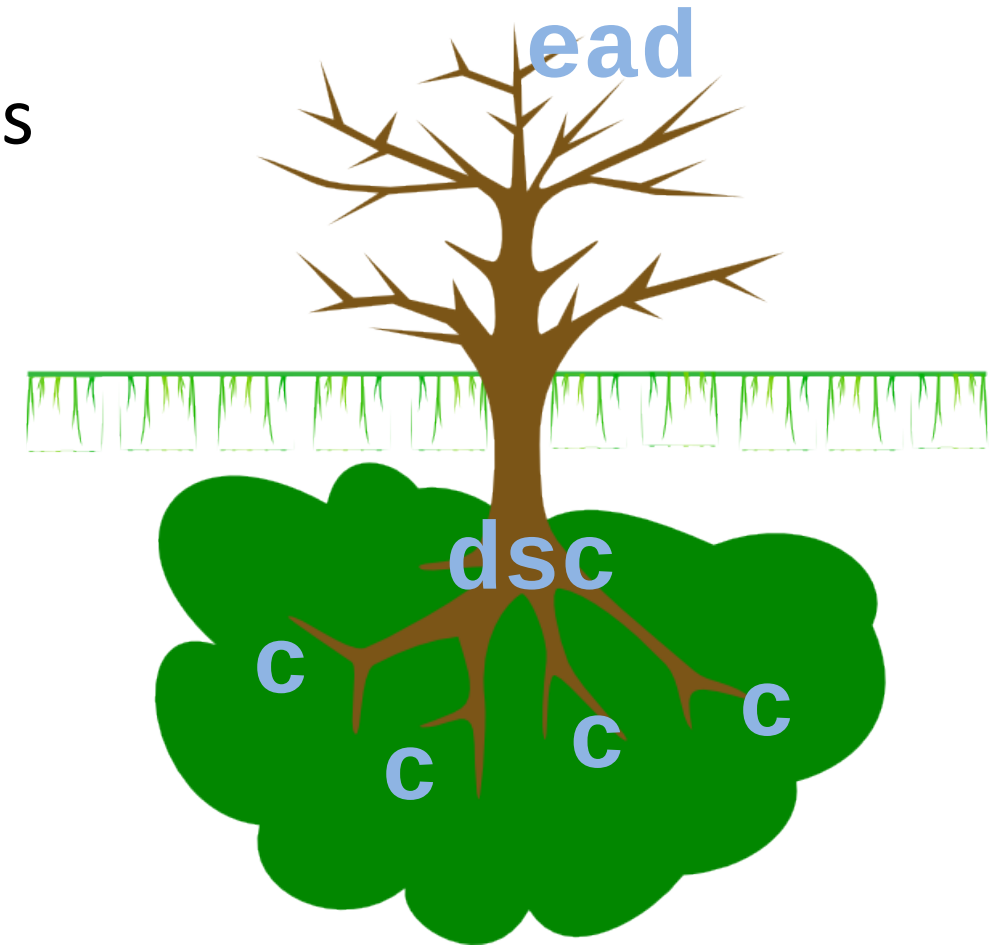
XTraining

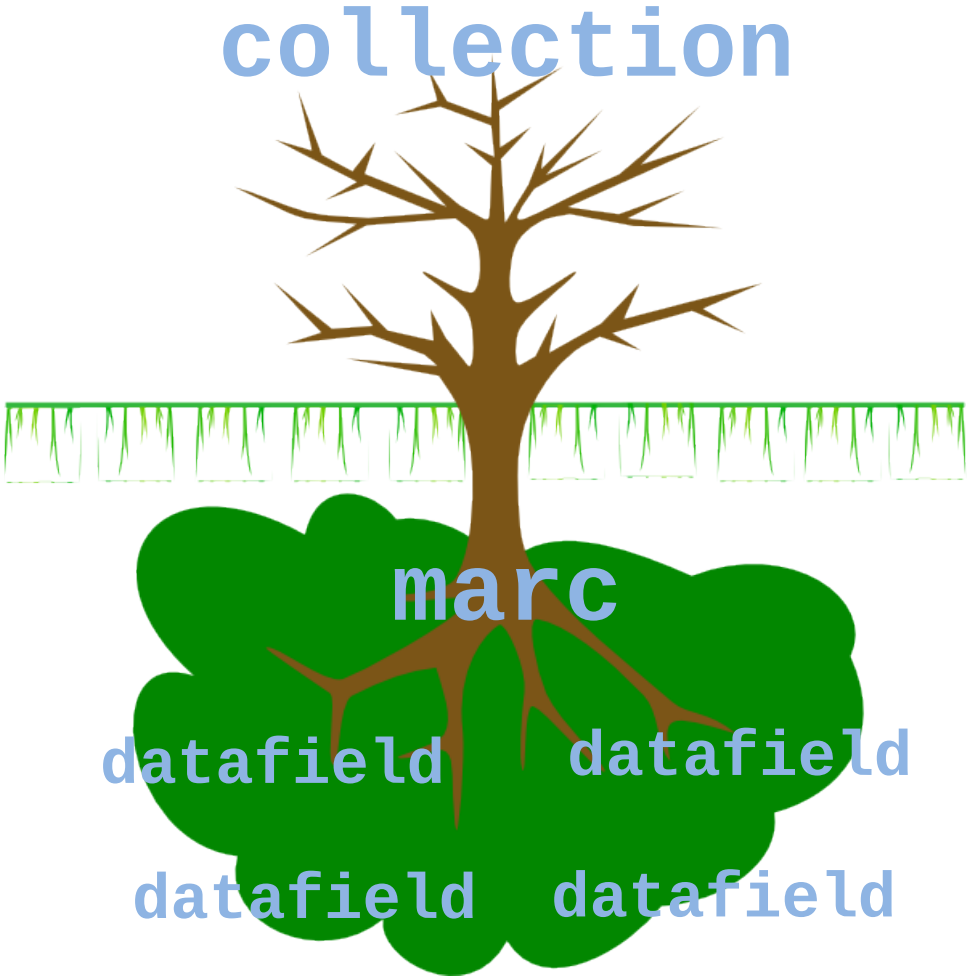
Part I: XPath

Regine Heberlein 5/9/2018

Helpful Concepts

- Well-formedness
- Root
- Node
- Tree



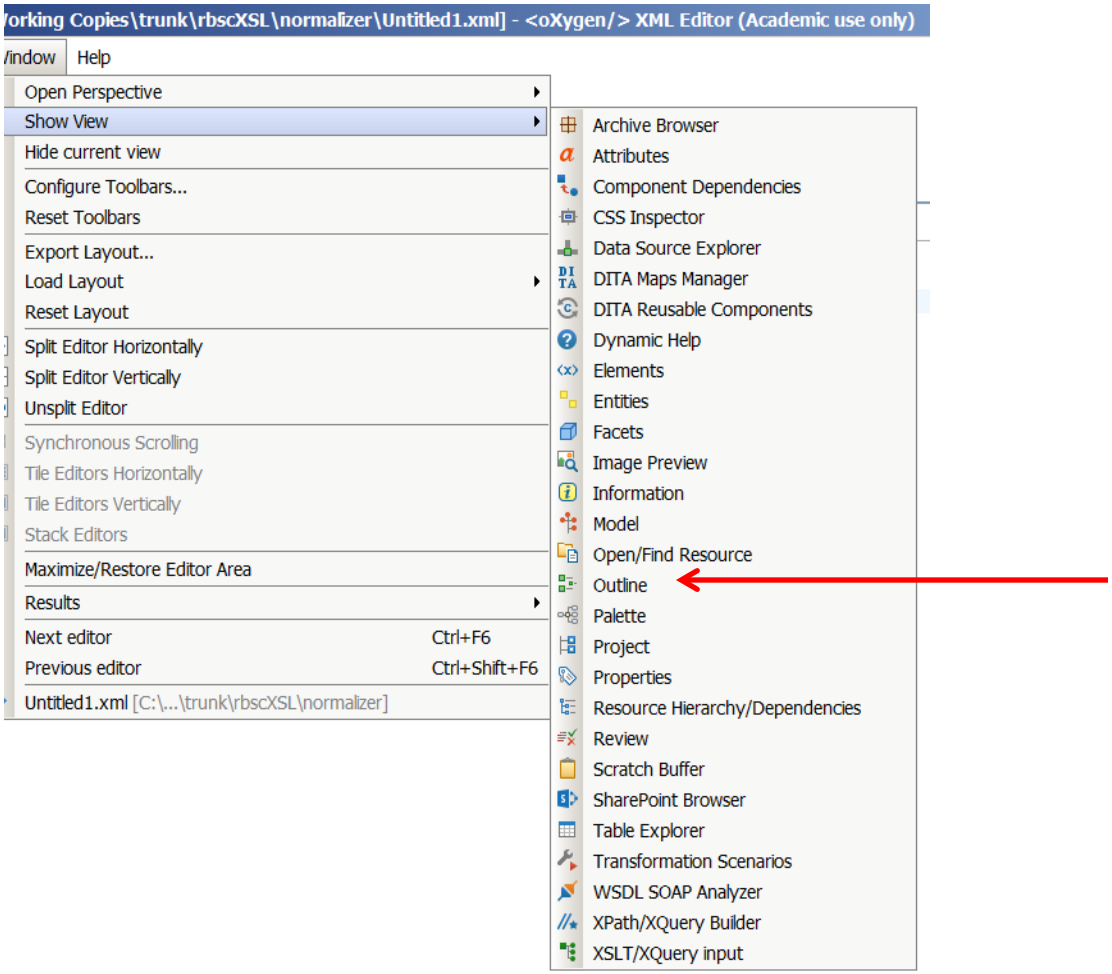

collection
marc
datafield
datafield
datafield
datafield

Node v. Element

Every element is a node,

but not every node is an element.

A Helpful oXygen Tool



How does XPath work?

An XPath expression may return

- a node-set OR
- a string OR
- a Boolean OR
- a number
- […]

First node tests

(don't forget to start with "//")

c

unittitle

unitdate

repository

address

*

@id

@altrender

@source

@calendar

@level

@*

text()

comment()

element()

attribute()

Context
(not the archival kind)

An XPath expression is evaluated

with respect to the context node

… this boat


Mirador #6
N Shoreline Blvd
Mirador #5
Mirador #4
Boat Yard
CC Yacht Club
Coopers Alley
Lawrence Street
Peoples Street
Bait Stand
Boaters Facility
Coopers Alley L Head
Dry Yard Stand
Public Boat Ramp
Fuel & Pump Out
Marina Office
Boaters Facility
Lawrence Street T Head
Restaurant
Yachting Center Restaurant
Restaurant
Boaters Facility
Peoples Street T Head

Stepping In (and Around)

//c/did
//c/did/unitdate
//c//unitdate
//c/did/unitdate/@normal
//unittitle/text()
//archdesc/descgrp/accessrestrict/p
//dsc//c/did/physdesc/extent
//archdesc/did//extent
//archdesc/controlaccess/persname
//controlaccess/persname

`Axis::node_test[predicate]`

e.g.,

parent::c[@level='file']

Axes Examples

child::item

parent::item

parent::item/text()

Axes: Verbose and Concise Syntax

Full Syntax	Abbreviated Syntax	Notes
ancestor		
ancestor-or-self		
attribute	@	@abc is short for attribute::abc
child		xyz is short for child::xyz
descendant		
descendant-or-self	//	// is short for /descendant-or-self::node()/
following		
following-sibling		
namespace		
parent	..	.. is short for parent::node()
preceding		
preceding-sibling		
self	.	. is short for self::node()

Try and discuss

- //c/child::did/child::container
- //c/descendant::container
- //extent/parent::physdesc
- //extent/ancestor::c
- //self::unittitle

What will these expressions return?

- //langusage/child::language
- //langmaterial/child::language
- //item/ancestor::revisiondesc
- //item/parent::revisiondesc

Let’s add some predicates

- Return any and all containers of type “box”

  //container[@type=‘box’]

- One step may have multiple predicates:

  //container[ancestor::c[1][@level=‘file’]][@type='box']

The order of predicates matters!

c[1][@level='file']

→matches if the first child c of the context node satisfies the condition @level='file'

c[@level='file'][1]

→matches the first child c of the context node that satisfies the condition @level='file'

Position tests

//c[1]

→the first c in context,

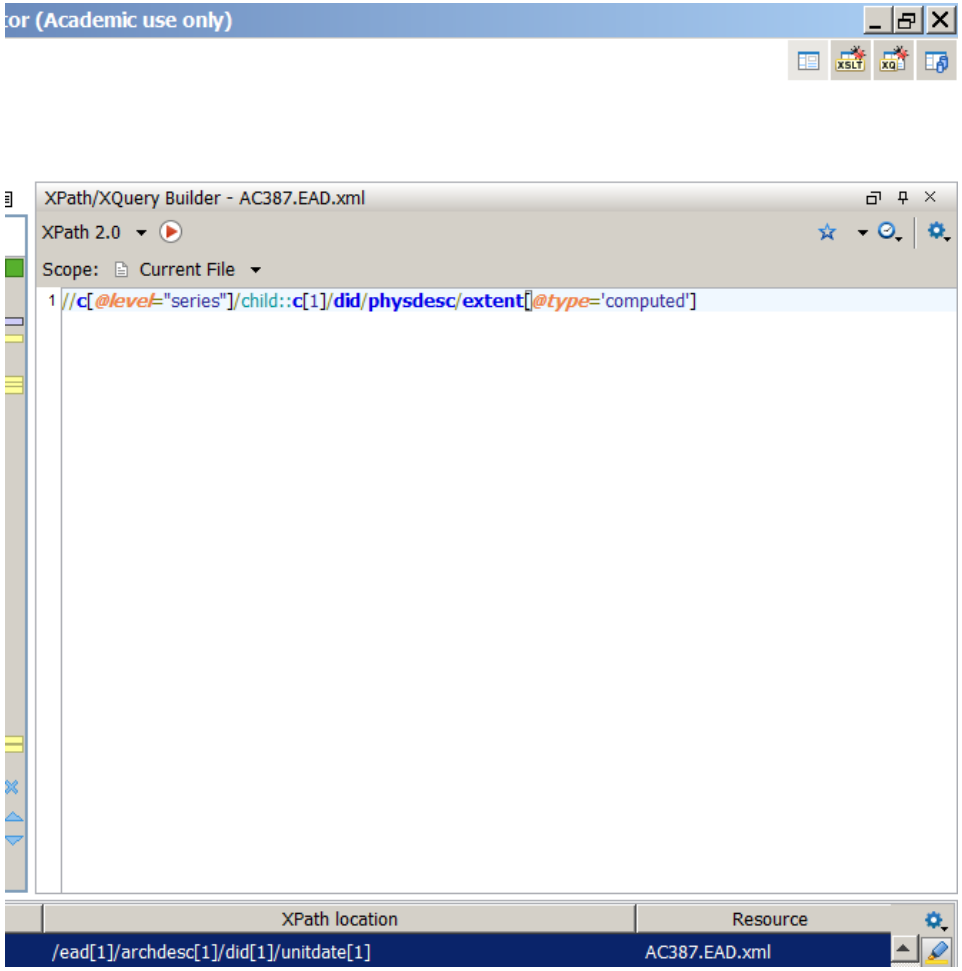
e.g. the first c of each series etc.

(//c)[1]

→ the first c in document order,

i.e. the absolute first c of the document

Another Useful Tool
Show View → XPath Builder



Position tests

- (//container)[1]
- (//container)[27]
- Return all elements named c whose level attribute is set to “file”

  //c[@level=”file”]
- Return the computed extent of the first c within each series in the document

  //c[@level="series"]/child::c[1]/did/physdesc/extent
  [@type='computed']
- Return the first inclusive unitdate in the document

  (//unitdate[@type=‘inclusive’])[1]

Challenge

Write an XPath to find:

1. All unittitle elements whose value equals “Malice in Wonderland” (hint: consider child elements and use predicate [.=“Malice in Wonderland”])
2. Any elements within a c whose value equals “Tartuffe” (hint: use a wildcard)
3. All emph elements whose value equals “Tartuffe” and that are descendants of the 36th c element.

Use Outline View to double-check your results!

- https://www.w3.org/TR/2017/REC-xpath-31-20170321/