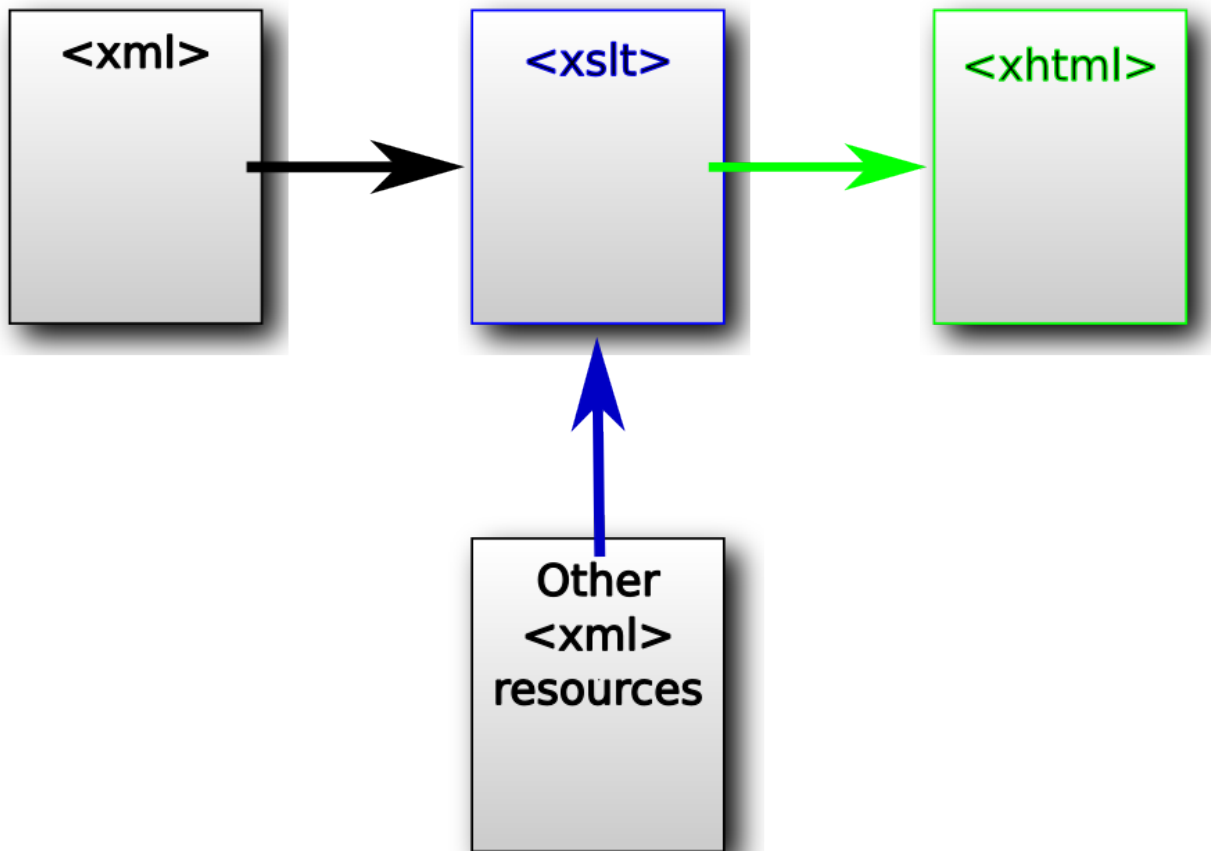


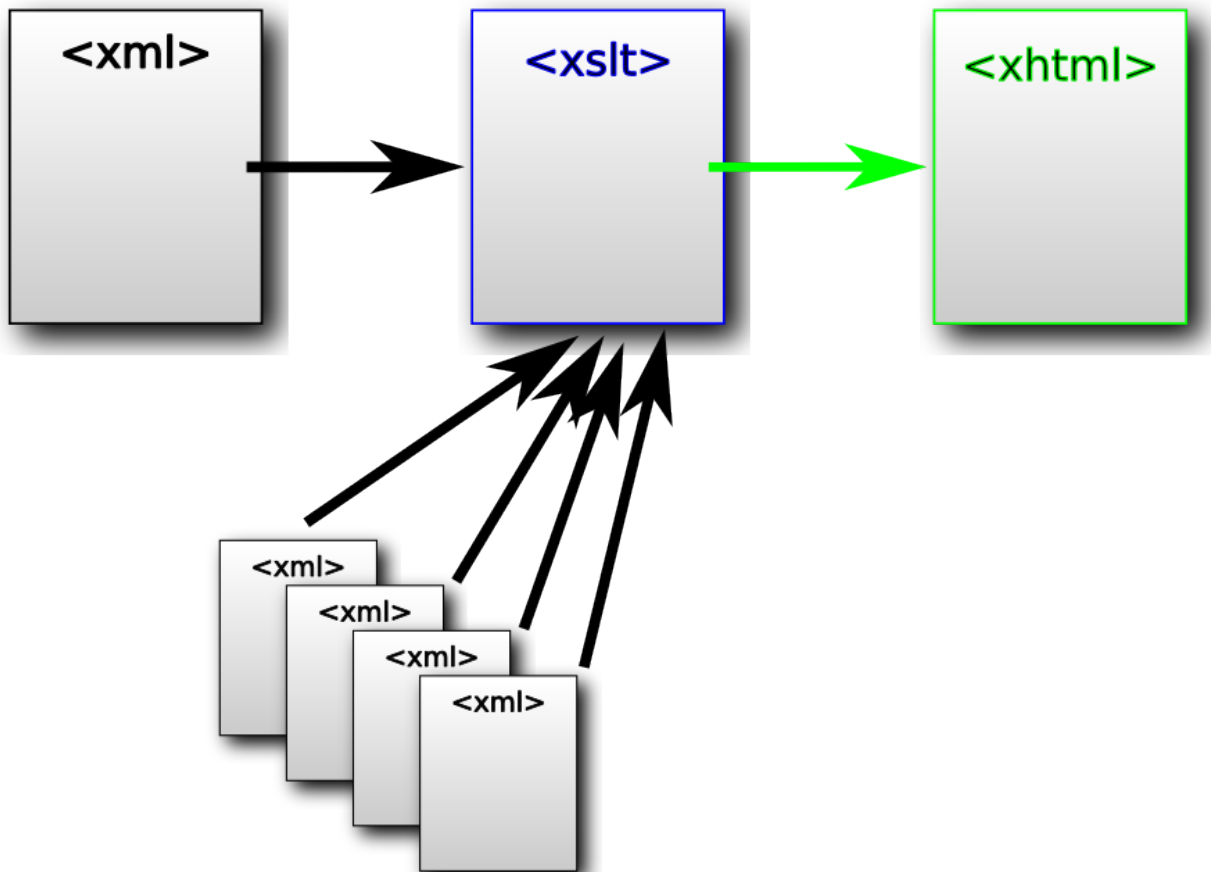
*The XPath collection() function:
pulling in multiple documents*

Martin Holmes

We know how to use doc()

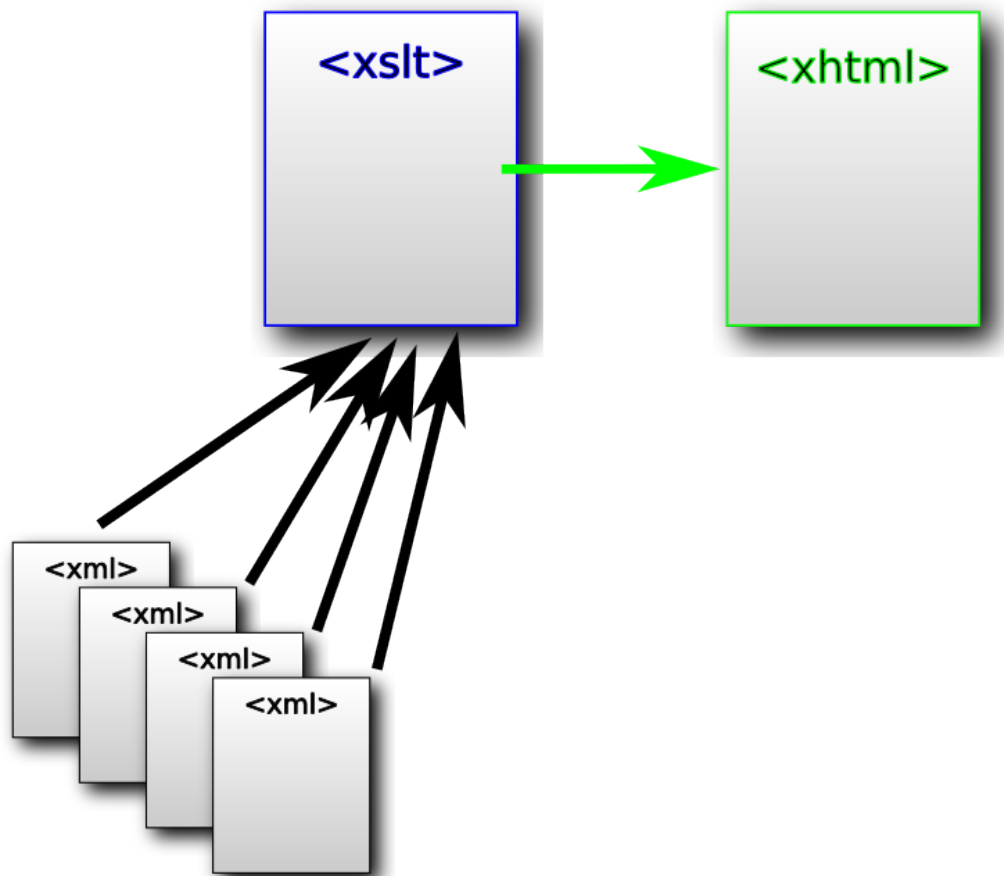


What if we want lots of documents?



We could laboriously load them all with `doc()`, of course, assuming we know their names.

Maybe you don't even have a main document



What kinds of scenario can we imagine for this? Mainly it's processing all the files in a collection or a subcollection.

The `collection()` function

The `collection()` function takes a URI parameter, which can point to files in three ways:

- A file containing a list of documents (a "catalogue").
- A folder containing documents.
- A folder along with a "filter" expression which limits the documents retrieved.

collection() pointing to a catalogue file

- Create a catalogue file called e.g. "catalogue.xml" which looks like this:

```
<collection stable="true">
  <doc href="chapters/chap1.xml" />
  <doc href="chapters/chap2.xml" />
  <doc href="chapters/chap3.xml" />
  <doc href="chapters/chap4.xml" />
</collection>
```

- Point to it like this:

```
<xsl:variable name="documents" select="collection('catalogue.xml')"/>
```

Remember all paths are relative. This assumes that the catalogue.xml file is in the same directory as the XSLT file. and all the other files are in a subdirectory called "chapters".

collection() pointing to a directory

```
<xsl:variable name="documents" select="collection('chapters')"/>
```

This will load all the documents in the directory called "chapters" which is a sibling of the XSLT file itself.

collection() pointing to a directory with a filter

```
<xsl:variable name="documents" select="collection('.*select=*.xml')"/>
```

This will load all the .xml documents in the directory containing the XSLT file itself.

A worked example

- Download the example file for collection(): <http://web.uvic.ca/~mholmes/dhoxss2013/examples/collection.xsl>
- Save it in the same directory as all your other example files.
- Make sure there are only XML and XSLT files in that directory.
- Now here's the trick: switch to the XSLT debugger, and set the XSLT file as both the XML and as the XSL file (in other words, it's processing itself).

Discuss this example thoroughly, going into what sequences are, and how we're now able to operate on the complete document collection as if it were a corpus.

collection() caveats

- `collection()` is "specified in very abstract terms, and it's likely that its detailed behaviour will vary considerably from one implementation to the next" Kay 726.
- It can be tricky to get your URIs right, especially with relative paths.