Working with zip files in XProc 3.0

GIT for this webinar: https://github.com/xatapult/markupuk-2020

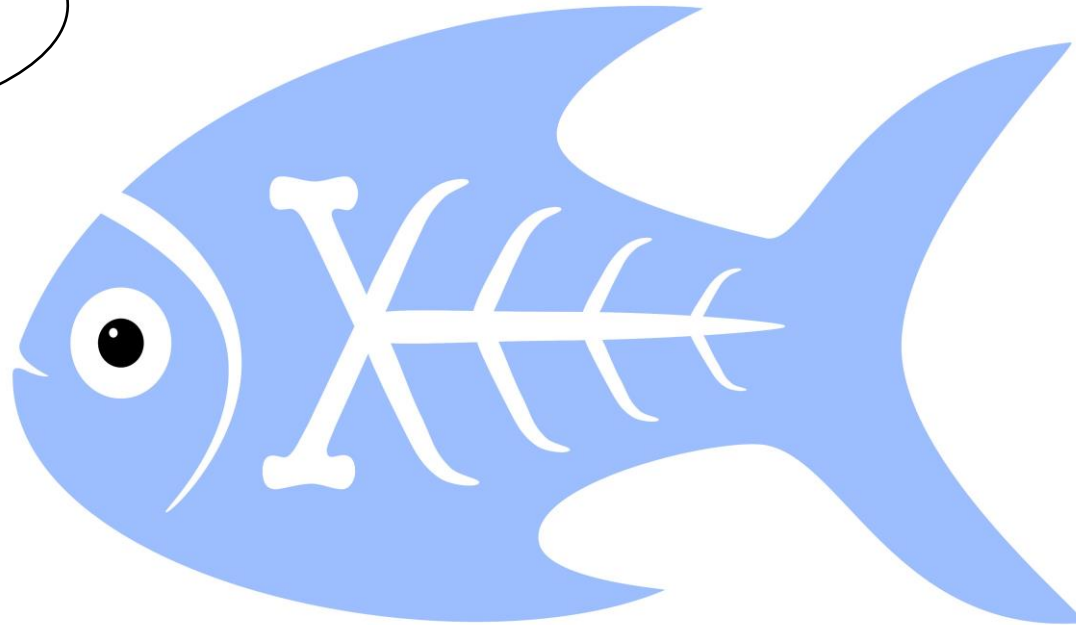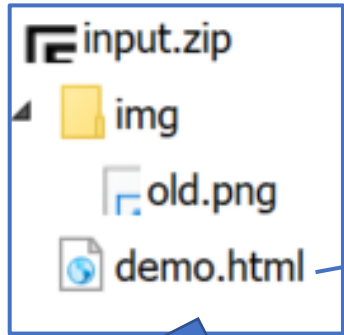# What's the input?

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>XProc demo file</title>
  </head>
  <body>
    <p>The old logo:</p>
    <p>
      <img src="img/old.png" alt="Old XProc logo"/>
    </p>
  </body>
</html>
```
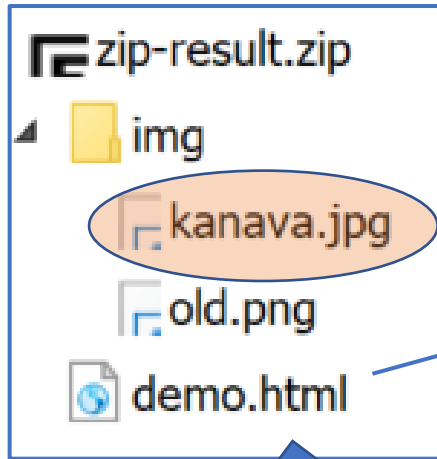
input.zip
  img
    old.png
  demo.html

Shows in the browser as…

The old logo:

What's the desired output?


```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>XProc demo file</title>
  </head>
  <body>
    <p>The old logo:</p>
    <p>
      <img src="img/old.png" alt="Old XProc logo" />
    </p>
    <p>New logo:</p>
    <p>
      <img src="img/kanava.jpg" alt="New XProc logo" width="10%"/>
    </p>
  </body>
</html>
```
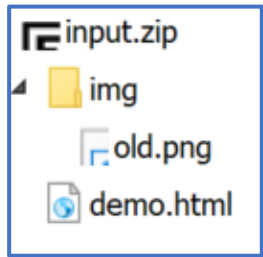
zip-result.zip
img
kanava.jpg
old.png
demo.html

Shows in the browser as...

The old logo:

New logo:

Of course using an XProc 3.0 pipeline…

URI of zip file   URI of file in zip

input.zip
▲ 📁 img
    old.png
📄 demo.html

Get manifest:
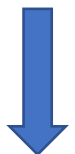<p:archive-manifest>

```
<c:archive xmlns:c="http://www.w3.org/ns/xproc-step">
    <c:entry name="demo.html"
             href="file:///.../input/input.zip/demo.html"
             method="deflated"
             size="219"
             compressed-size="150"
             time="2020-02-06T08:19:06+01:00"/>
    <c:entry name="img/old.png"
             href="file:///.../input/input.zip/img/old.png"
             method="deflated"
             size="7522"
             compressed-size="7043"
             time="2019-03-14T15:22:48+01:00"/>
</c:archive>
```

See: **markupuk-2020/zip-handling/example/zip-add-new-logo-1.xpl**

input.zip
- img
  - old.png
- demo.html

Get manifest

Add reference to new logo to manifest
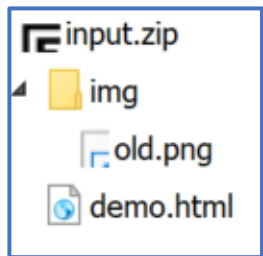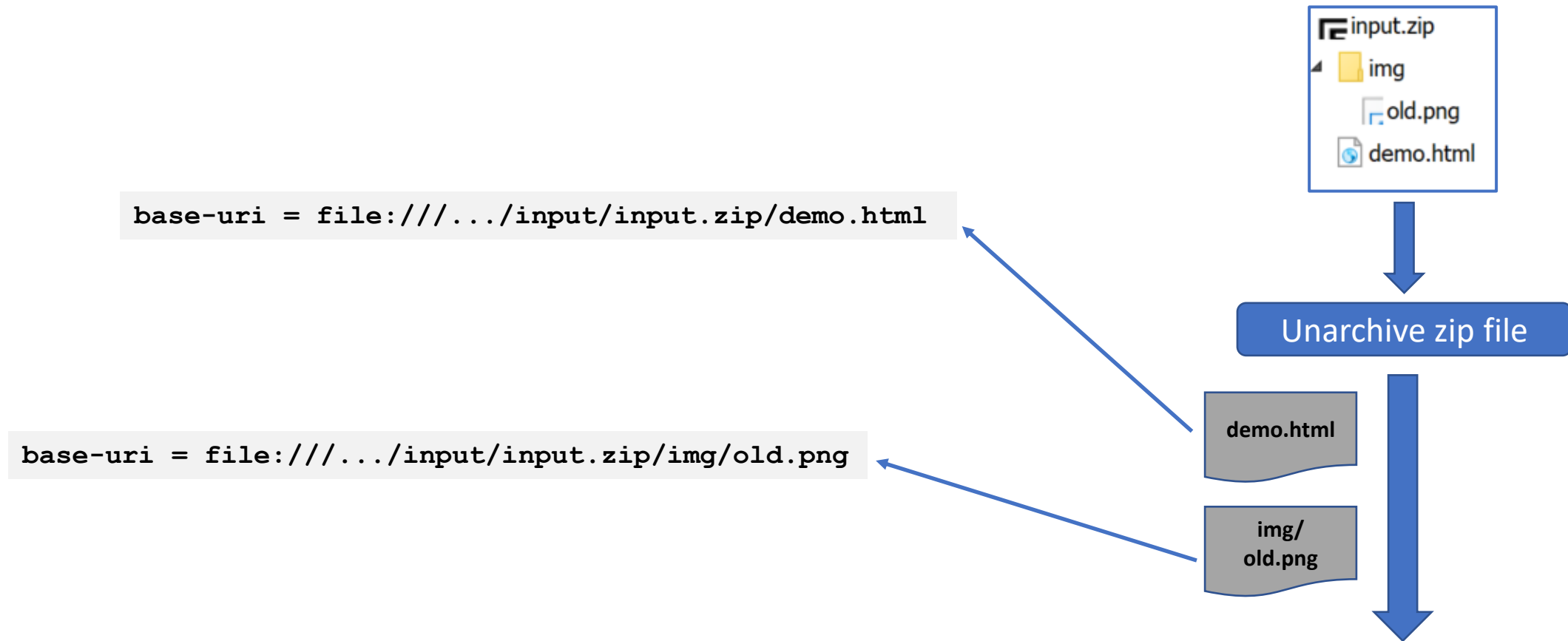
```
<c:archive xmlns:c="http://www.w3.org/ns/xproc-step">
    <c:entry name="demo.html"
             href="file:///.../input/input.zip/demo.html"
             method="deflated"
             size="219"
             compressed-size="150"
             time="2020-02-06T08:19:06+01:00"/>
    <c:entry name="img/old.png"
             href="file:///.../input/input.zip/img/old.png"
             method="deflated"
             size="7522"
             compressed-size="7043"
             time="2019-03-14T15:22:48+01:00"/>
    <c:entry name="img/kanava.jpg"
             href="file:/.../additional-images/kanava.jpg"/>
</c:archive>
```

URI of file on disk

See: **markupuk-2020/zip-handling/example/zip-add-new-logo-2.xpl**

`base-uri = file:///.../input/input.zip/demo.html`
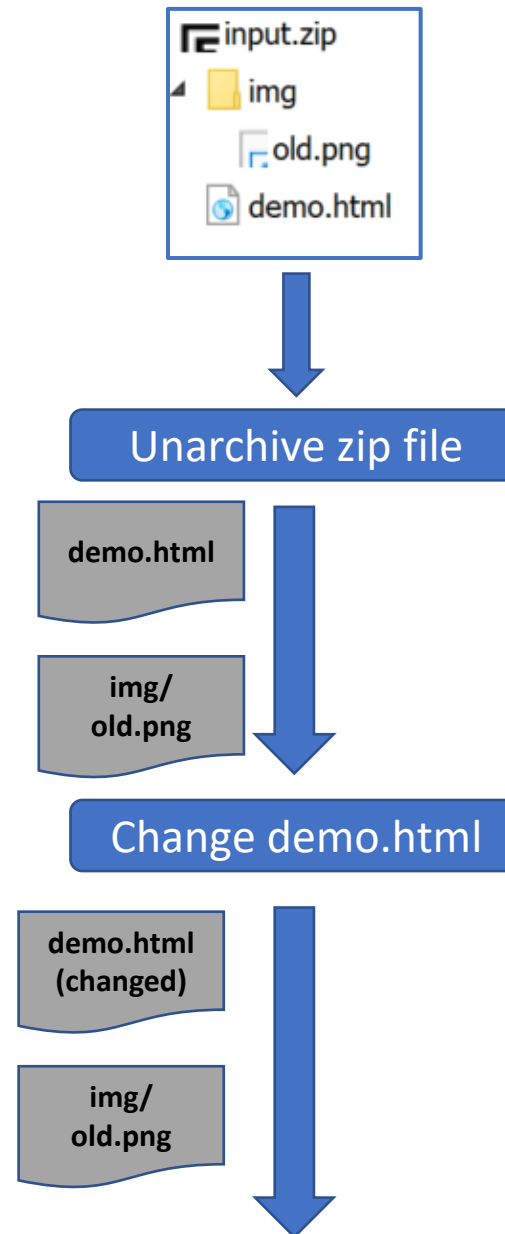
`base-uri = file:///.../input/input.zip/img/old.png`

All 2 documents in the archive with property **base-uri** same as **@href** in the manifest!

See: **markupuk-2020/zip-handling/example/zip-add-new-logo-3.xpl**

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>XProc demo file</title>
  </head>
  <body>
    <p>The old logo:</p>
    <p>
      <img src="img/old.png" alt="Old XProc logo" />
    </p>
    <p>New logo:</p>
    <p>
      <img src="img/kanava.jpg" alt="New XProc logo" width="10%"/>
    </p>
  </body>
</html>
```
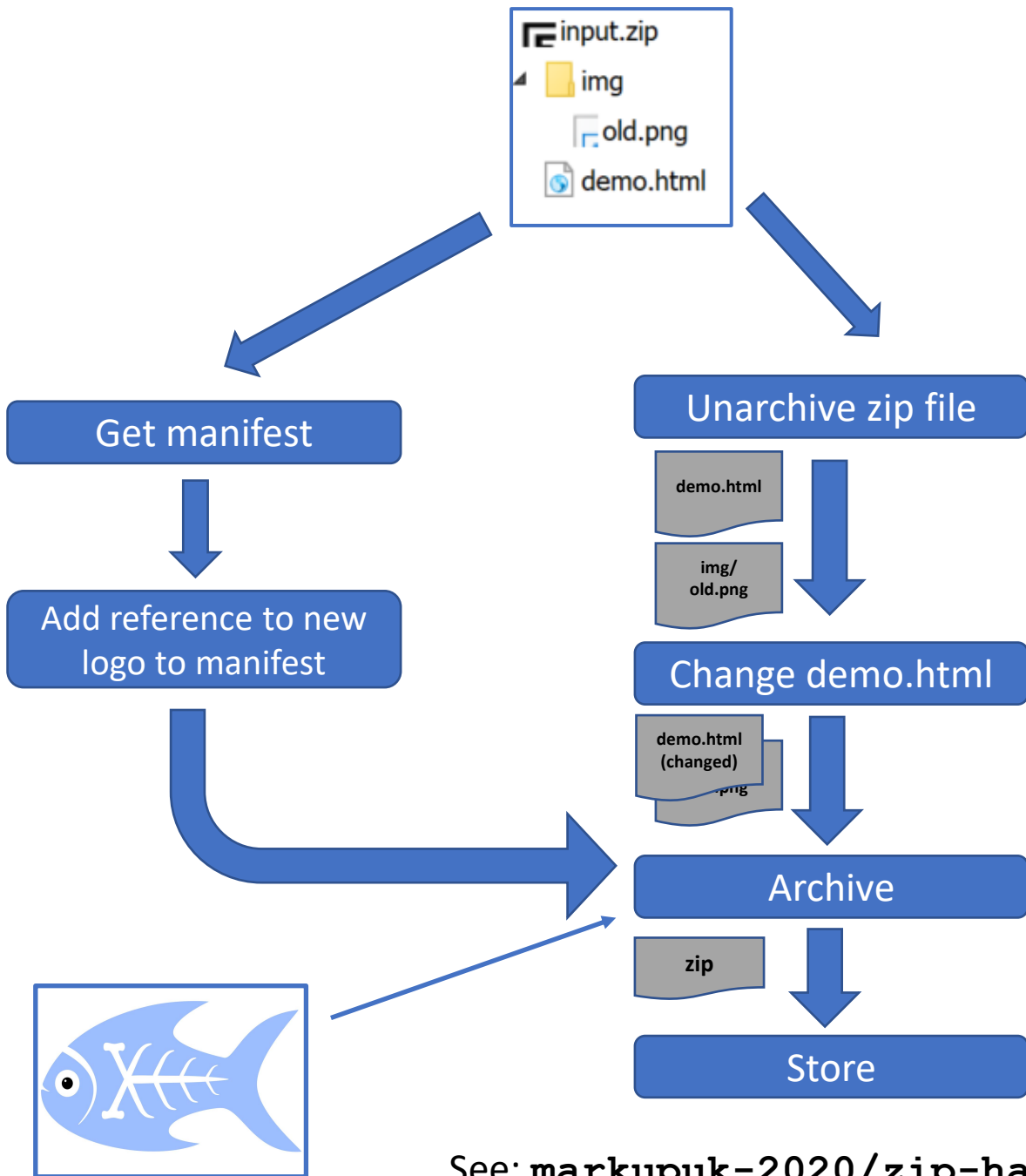
No changes in **base-uri** properties!

input.zip
  img
    old.png
  demo.html

Unarchive zip file

demo.html

img/
old.png

Change demo.html

demo.html
(changed)

img/
old.png

See: **markupuk-2020/zip-handling/example/zip-add-new-logo-4.xpl**

```
<c:archive xmlns:c="http://www.w3.org/ns/xproc-step">
    <c:entry name="demo.html"
             href="file:///.../input/input.zip/demo.html"
             method="deflated"
             size="219"
             compressed-size="150"
             time="2020-02-06T08:19:06+01:00"/>
    <c:entry name="img/old.png"
             href="file:///.../input/input.zip/img/old.png"
             method="deflated"
             size="7522"
             compressed-size="7043"
             time="2019-03-14T15:22:48+01:00"/>
    <c:entry name="img/kanava.jpg"
             href="file:/.../additional-images/kanava.jpg"/>
</c:archive>
```

- Match documents in the manifest with documents on the source port, by `base-uri` property
- If not, try to load from URI in manifest @`href`

See: `markupuk-2020/zip-handling/example/zip-add-new-logo-5.xpl`

Wrap up:

- There are 3 steps for workin g with archives
  - `<p:archive-manifest>`
  - `<p:unarchive>`
  - `<p:archive>`
- Archive manifest:
  - For finding out what's in an archive
  - For creating an archive
- Select files for the archiving using document's `base-uri` values
  - Either from documents on the `source` port or on disk

Goodbye and thank the fish, again!

Your guide today: Erik Siegel – erik@xatapult.nl

Specification: https://spec.xproc.org/

Processors:

- Morgana: https://www.xml-project.com/
- Calabash: https://xmlcalabash.com/

Articles on XProc: https://www.xml.com

Book: https://xmlpress.net/publications/xproc-3-0/

See you!
And remember,
Kanava says:
XProc rocks...

Who Am I?

- Erik Siegel
- Content Engineer, XML Specialist, Technical Writer
- Company: Xatapult
  - Groningen, The Netherlands
  - Customers mostly in publishing and standardization
- Part of the XProc 3.0 editing committee
- Writer of the XProc 3.0 Programmer Reference
- Contact:

  erik@xatapult.nl
  www.xatapult.com
  www.linkedin.com/in/esiegel/
  +31 6 53260792