Using File URIs

Since Fedora 3.3 it is possible to reference managed and externally-managed content (type "M" and "E") with a `file:` URI within the digital object for the ingest. In order to enable this functionality the following changes are necessary:

First edit the predefined XACML policy, uncomment the relevant rule in the preinstalled policy file `deny-unallowed-file-resolution.xml` and adapt the regex to your needs (or write your own policy). Optionally bind the rule to a specific user. Make sure the RuleId is unique.

```
...
<Rule RuleId="1" Effect="Permit">
  <Condition FunctionId="urn:oasis:names:tc:xacml:1.0:function:regexp-string-match">
    <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">{^}{{[file:/allowed/.*$]}}<
/AttributeValue>
    <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-one-and-only">
      <ResourceAttributeDesignator AttributeId="urn:fedora:names:fedora:2.1:resource:datastream:fileUri"
DataType="http://www.w3.org/2001/XMLSchema#string"/>
    </Apply>
  </Condition>
</Rule>
...
```

Then create your digital object FOXML and use the `file:` URI where necessary.

```
...
  <foxml:datastream CONTROL_GROUP="E" ID="MEDIUM_SIZE" STATE="A" VERSIONABLE="true">
    <foxml:datastreamVersion CREATED="2008-07-02T05:09:42.937Z"
        ID="MediumSize.jpg.0" LABEL="Medium-size image" MIMETYPE="image/jpeg">
      <foxml:contentLocation REF="file:///path/to/files/image.jpeg" TYPE="URL"/>
    </foxml:datastreamVersion>
  </foxml:datastream>
...
```

file URI format

The provided `file:` URIs must not have an authority component or the authority component must be empty. This means that URIs with the `file:` scheme must either have one slash (no authority component) after the scheme:
`file:/data/image.jpeg`
or have three or more slashes (an empty authority component) after the scheme:
`file:///data/image.jpeg`
If you use this form the regex in the policy that matches the "one slash form" will nontheless match because internally Fedora translates all `file:` URIs into the one slash form.
As a result {^}`file:/data/.*$` will match both forms above.

The use of two slashes after the scheme is not allowed and will result in an error because it defines an authority component.

File URIs and externally-managed datastreams

The default policy `deny-unallowed-file-resolution` only allows authenticated users to retrieve files from the allowed file paths. For managed content (type "M"), this restriction only applies at datastream creation; once the datastream is created, it is available via the API-A methods to any user allowed by your policies (by default, API-A methods are unrestricted). However, externally-managed content is protected by the `deny-unallowed-file-resolution` policy at the moment the datatream is created, and thereafter every time the datastream is accessed via either API-A and API-M methods; please be aware that this has the effect of making externally-managed content with file URIs available only to authenticated users, such as `fedoraAdmin`.

Note:

Please bear in mind that the activation of `file:` URIs for managed content exposes your filesystem to the ingest process and as such could be abused by inserting URIs to files that are not intended for ingestion. While Fedora sanitizes the given URI and denies URIs such as file:///data/../etc/passwd, Make sure that you:

- only expose directories without symlinks
- only expose directories that don't contain any sensitive information, like access to configuration files, password files, user home directories, etc.
- deny file URIs as soon as the ingest is finished