

Fedora Identifiers

PIDs

A PID is a unique, persistent identifier for a Fedora digital object. PIDs may be user-defined or automatically assigned by a repository. In this section we describe the syntactic and normalization considerations for PIDs.

Syntax

PIDs are case-sensitive and consist of a namespace prefix and a simple string identifier. The syntax is described below using [augmented BNF](#):

```
object-pid      = namespace-id ":" object-id
namespace-id    = 1*( ALPHA / DIGIT / "-" / "." )
object-id       = 1*( ALPHA / DIGIT / "-" / "." / "~" / "_" / escaped-octet )
escaped-octet   = "%" HEXDIG HEXDIG
```

The maximum length of a PID is 64 characters.

For convenience, we provide the following single regular expression, which can be used to validate a normalized PID string:

```
^[A-Za-z0-9](-|\.){0,63}((A-Za-z0-9)(-|\.|\~|_|(%[0-9A-F]{2})){0,63})+$
```

Examples

- demo:1
- demo:A-B.C_D%3AE
- demo:MyFedoraDigitalObject

Normalization

HEXDIG characters may occur in lowercase, but should be capitalized for normalization purposes. The separator character may occur as "%3A" or "%3a", but should be changed to a colon ":" for normalization purposes.

Datastream IDs

Datastreams IDs may consist only of [XML NCName characters](#) and must not exceed 64 characters in length.

URIs for Objects

It is often useful to have Uniform Resource Identifiers ("URIs") that refer to Fedora Objects. For instance, semantic web technologies require the use of a URI to identify a subject. Other benefits of exposing and using URIs are described in [Section 2 of the W3C's *Architecture of the World Wide Web*](#).

Every Fedora object has an implicit URI associated with it. These identifiers exist within the "fedora" namespace of the "info" URI scheme. We chose this URI scheme due to its resolution protocol independence and syntactic freedom.

Syntax

The URI for a Fedora object is constructed simply by appending the PID to the string "info:fedora/".

Examples

- info:fedora/demo:1
- info:fedora/demo:A-B.C_D%3AE
- info:fedora/demo:MyFedoraDigitalObject

Normalization

To normalize an object URI, normalize the PID part as described above.

URIs for Disseminations

Every dissemination of an object also has an implicit URI associated with it. This is useful when describing or referring to the representations provided by a digital object.

Syntax

Dissemination URIs take one of two forms. In the case of a method call the URI indicates the service definition and the method (along with any parameters). In the case of a datastream dissemination, the URI indicates the Datastream id.

```
dissemination-uri = "info:fedora/" pid "/" ( method-call / datastream-id )
method-call       = sDef-pid "/" method-name [ "?" param *( "&" param ) ]
param             = paramName "=" paramValue
```

Note: Although datastream-ids and method-names may consist of XML NCName characters. NCName characters that are not URI-safe must be escaped using one to four escaped UTF-8 octets per character, each of the form "%" HEXDIG HEXDIG.

Examples

- info:fedora/demo:1/demo:MySDef/method
- info:fedora/demo:1/demo:MySDef/method?param1=value1
- info:fedora/demo:1/title.jpg
- info:fedora/demo:1/DC

Normalization

To normalize a dissemination URI:

1. Normalize the PID portion(s) of the URI.
2. Un-escape any URI-escaped characters that do not need escaping according to the definition of the "info" scheme.
3. Make all remaining escaped octets use UPPERCASE (%ff becomes %FF).
4. Parameters should be alphabetized in order by name, then by value. The order should be according to occurrence in UTF-8.