discussion of goals

For Identifiers (just the complete URI and local part, not about mechanisms)

Permanent

The PID system should provide long term support for resolving identifiers. The hope is that groups at Cornell and outside of Cornell will be able to use these identifiers to signify digital items, concepts, and other things in a stable maner. The reason to pursue this goal is to reduce the cost of creating systems at Cornell that can refer to things created by other systems.

Permanently resolve to a digital object

The Identifiers should be permanently resolvable via HTTP to a digital resource.

A large category of items that a PID resolver at CUL would index would be digital objects that should remain fixed but it is not a goal of the PID resolver system to ensure fixity of digital resources. There are also items that are outside of this category that the PID resolver should also work with.

We may want to have identifiers for digital objects that constantly change. We may want to have identifiers for digital objects that represent the state of non-digital resources.

URL

The PIDs should conform to the URL RFC1738

Resolvable through a web browser

Since the PIDs are URLs they could be used as the href attribute in an <a> element in HTML. When requested via HTTP GET the response should be a digital object or digital surrogate. Wrinkle in this goal is that many web browsers do not respect the MIME type of responses. A digital object such as a PDF might need to have the part after the final slash end if .pdf to be interpreted as a PDF file by a browser. Some browsers will attempt to open a result as a PDF file on a GET of a URL like http://resolver.org/234/b233.pdf that has a reply status of 400.

If the ID schema has a DNS name then all PIDs in the system should be unique even with the DNS name removed. This goal is trivially satisfied if there is only one DNS name used by the PID system. If there are more than one DNS names used in the system care must be taken to ensure that goal is met.

Works with existing systems

The PIDS should work with existing systems such as VIVO, arXiv, OAIS (CUL), Voyager Catalog, and WorldCat etc.

Simplicity

Please add comments about this goal.

Support for a local namespace prefix and an identifier part

The hope is that the URL can be of the form http://hostname.com/localNamespacePrefix/identifierPart or something similar. The localNamespacePrefix could also be called a collectionPrefix. The resolver system should place no additional restrictions on the identifierPart beyond conforming URL syntax.

Support for opaque identifiers

In an attempt to avoid problems in situations where the labels associated with resources change, the PIDs should support partially opaque identifiers such as http://hostname.com/170/2a33-ffff instead of http://hostname.com/SuperMegaCollection/WalterCarlos1. Since the resolver system places no additional restrictions on the identifier part, we cannot stop systems from requesting new identifiers with a syntax that they attach meaning to.

Ex. A collection administrator might register the following Identifiers: http://resolver.cornell.edu/170/article23332 -> http://collectionX.cornell.edu/article/23332 http://resolver.cornell.edu/170/article23332.pdf -> http://collectionX.cornell.edu/article/23332?format=pdf http://resolver.cornell.edu/170/article23332.tex -> http://collectionX.cornell.edu/article/23332?format=tex
The resolver system will not attempt to parse these identifiers and will not record or track relationships between identifiers.

Local part should not be Cornell branded

To encourage the possibility of using the PIDs across intuitions, the PIDs should not be branded.

May be surrogate for physical object

The digital resource returned by the PID resolver might not be the thing identified by the PID, but a surrogate for a resource that cannot be transported easily via HTTP.

Should be short

Please add comments about this goal. How short? Should there be a length limit (which implies a limit to the number of things that can be resolved)?

Should be easy to copy by hand

Please add comments about this goal. Adding dashes between every 4 digits of the identifier may be a way to improve the easy of copying the PIDs.

For Resolver and System

Supports billions of identifiers

Please add comments about this goal.

Robust architecture

Please add comments about this goal.

Robust implementation

Please add comments about this goal.

Ability to request metadata about the identifier

Please add comments about this goal.

Lightweight understanding of identifier equivalence

Please add comments about this goal.

Should be easily discoverable by Google

Please add comments about this goal.

Integrates well with the web architecture

This should be satisfied by earlier goals.

Vitality checking

Please add comments about this goal.

Need to avoid unbounded generation of surrogate persistent identifiers

Please add comments about this goal.

PID corresponding to every Cornell NetID

Please add comments about this goal.

Useful error when check digit is wrong

Please add comments about this goal.

Governance Issues

Requires payment to external organization

Please add comments about this goal.

Can continue to resolve IDs in the absence of the external organization

Please add comments about this goal.