pilin persistent identifier linking infrastructure

web: http://resolver.net.au/hdl/102.100.272/0N8J991QH
email: policy@pilin.net.au

Version History

Version	Date	Status & changes	Expression identifiers
V1.0	2007-09-28	Initial release for feedback	PILIN/84L8V1FQH hdl:102.100.272/84L8V1FQH
V2.0	2007-11-26	Convert from table to text	PILIN/ QD9QVGMQH Hdl:102.100.272/QD9QVGMQH

Identifier Policy FAQ

To cite the latest version of this work use http://resolver.net.au/hdl/102.100.272/PFKJT1FQH
To cite this version of this work, use http://resolver.net.au/hdl/102.100.272/QD9QVGMQH

1 Scope

This document answers questions about identifiers. Reference is made to the community guideline and consideration documents developed by the PILIN team as well as the PILIN Policy Documents – (what the PILIN team used during the course of the project) – some questions have reference to both types of documents

2 Questions

Should labels be meaningful?

Is it OK to have meaningful identifiers?

How long should labels be?

What should labels look like?

Copyright © Monash University

This work is licensed under the Creative Commons Attribution-Share Alike 2.5 Australia License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/2.5/au/

This work was created as part of the PILIN project. The PILIN project is funded by the Australian Commonwealth Department of Education, Science and Training, (DEST) under the Systemic Infrastructure Initiative (SII) as part of the Commonwealth Government's Backing Australia's Ability – An Innovation Action Plan for the Future (BAA) under the ARROW Project.

How should Identifiers be cited in documents?

What to identify?

When to identify?

How should an identifier system be set up?

What are the possible arrangements for hosting naming authorities?

How should appropriate Names for Contexts be chosen?

How should I manage multiple contexts?

What authority metadata should be associated with any identifier created?

Can I keep my existing identifiers if I move them to a new identifier scheme?

Do I give identifiers to different versions of an object?

Do I give identifiers to different presentation?

Do I give identifiers to different copies?

How in Handles do I relocate my naming authority to a different institution?

Do I have a single resolution service for my identifiers?

How do I make sure an Identifier is persistent while I have control over it?

How do I make sure an identifier is persistent if I no longer have control over it?

How do I get my own Handles?

How do I ensure that a service operating on a persistent identifier (eg a resolver) is itself persistent?

How can I use types or service identifiers with handles to ensure that handle identifiers is are actioned in a specific way?

Shouldn't I just stick to URLs for my identifiers?

How do I decide what my identifier labels should look like?

Should labels be meaningful?

Is it OK to have meaningful identifiers?

Community Guidelines and Considerations

- Semantically transparent labels are easy to generate and remember, and allow metadata to be worked out by inspection. But they do not persist when the thing they identify changes.
- Obfuscated identifiers are less vulnerable to change, and still allow metadata to be extracted. But expectations can still fail when thing changes.
- Meaningful identifiers can be treated as arbitrary by policy, with no guarantee that they will stay meaningful. This guarantees their persistence, but makes them much more confusing for users, and harder to police.

Meaningfulness of Labels in Identifiers
http://resolver.net.au/hdl/102.100.272/D6N8F0DQH

PILIN Policy – what we used in the PILIN Project

Semantically opaque. That is, no meaning is deliberately encoded into the suffixes. The scheme will also avoid accidentally creating potentially meaningful suffixes where possible.

Handle Suffix Format Policy for the PILIN Project http://resolver.net.au/hdl/102.100.272/VY0KYS0QH

How long should labels be?

What should labels look like?

Community Guidelines and Considerations

- Labels on URL should be URL-safe
- Arbitrary Labels should exclude punctuation
- Labels should not require preprocessing to a canonical form
- Labels should be short
- Labels should be large enough
- Labels should be uncased
- Arbitrary Labels should avoid confusable characters
- Arbitrary Labels should have the same length
- Label generation should avoid collisions

Form of Labels
http://resolver.net.au/hdl/102.100.272/0HJ9X8JQH

PILIN Policy – what we used in the PILIN Project

1. Easily citable and succinct. That is, the scheme will be short enough to allow transcription by hand and oral dictation.
2. Large enough to identify at least 10^9 different things.

Handle Suffix Format Policy for the PILIN Project
http://resolver.net.au/hdl/102.100.272/Y35XS0QH

How should Identifiers be cited in documents?

Community Guidelines and Considerations

- An identifier is not the same thing as a service call to resolve the identifier. (The difference is more obvious in schemes like Handle than in PURL, where the identifier is presented as a resolvable URL.)
- Different service calls have their own encodings, which may differ from the encoding of the identifier in isolation.
- A service call to resolve a persistent identifier is not necessarily as persistent as the identifier itself. The resolution service is not necessarily managed by the same party as the identifier.

Identifier Services Guidelines
http://resolver.net.au/hdl/102.100.272/1KKBLPDQH

PILIN Policy – what we used in the PILIN Project

The citation should be:

1. Resolvable. That is, the citation invokes a service that resolves the identifier.
2. Actionable in common web browsers. That is, the citation should be 'clickable' in common web browsers.
3. Persistent. That is, the citation should provide reliable action over time.
4. Respectful of the identifier owner's own citation policy.

Balancing these requirements sometimes necessitates the use of multiple citation encodings.

Citation of Handles within PILIN Documentation ,
http://resolver.net.au/hdl/102.100.272/R67T0T0QH

What to identify?

Community Guidelines and Considerations

- Have an information model. The things which can have an identifier include: Versions; Presentations (e.g. formats); Copies; Aggregates; Disaggregations; Service Transformations.
- Aggregations and Disaggregations are modelled if they are meaningful for business processes.
- Work out how identifiers will be managed. How you manage identifiers should not be tightly bound to you manage your information.
- Work out how identifiers will be used. (e.g. what services will interact with the identifiers)
- Define a notion of curation boundary (e.g. release)
- Persistently identify things that are:
- Published
    - Under your control
    - Stable
    - Conceptually meaningful (e.g. meaningful aggregations)
    - Citable
    - Describable through metadata.

Identifier Association Guidelines
http://resolver.net.au/hdl/102.100.272/WBNMH9DQH

PILIN Policy – what we used in the PILIN Project

The PILIN Project assigns an identifier to the following things from the PILIN information model[1]:

[1] The PILIN information model is described in *Identifier Association Policy for the PILIN Project.*

- every *citable work* that crosses the PILIN curation boundary,
- every *citable expression* of a citable work (e.g. version, draft, translation), and
- every *citable manifestation* of a release that crosses the PILIN curation boundary (e.g. document format, file format).

Identifier Association Policy for the PILIN Project
http://resolver.net.au/hdl/102.100.272/3SYFCJ3QH

When to identify?

Community Guidelines and Considerations

Associate an identifier with a thing:

- After you have decided to move the thing across the curation boundary
- Before the thing moves across the curation boundary
- Before the identifier moves across its own curation boundary.

Identifier Association Guidelines
http://resolver.net.au/hdl/102.100.272/WBNMH9DQH

PILIN Policy – what we used in the PILIN Project

The PILIN Project shall assign an identifier to a thing:

- after the decision has been made to move the thing across the PILIN curation boundary (separating who has write-access from who has read-access only); and
- before the thing actually moves across the PILIN curation boundary.

Identifier Association Policy for the PILIN Project
http://resolver.net.au/hdl/102.100.272/3SYFCJ3QH

How should an identifier system be set up?

What are the possible arrangements for hosting naming authorities?

Community Guidelines and Considerations

- You can either host your own system, or use someone else's. If you host your own, you have full control, but also full responsibility and management workload.
- You can have either exclusive access to the system, or share it with others. If you have exclusive control, you have clear governance, full control, and branding; but you cannot move identifiers to others easily, and cannot share out the management load.

Depending on those questions, hosting can be **devolved** (own system, exclusive access); **centralised** (external system, shared access); **autonomous** (external system, exclusive access: everyone gets their own system, but it is hosted centrally); or **federated** (own system, exclusive access: the federation owns the separate systems).

Considerations for Ownership of Identifier Management Systems
http://resolver.net.au/hdl/102.100.272/461BL3DQH

How should appropriate Names for Contexts be chosen?

Community Guidelines and Considerations

Context names are usually semantically transparent, which allows straightforward branding and discovery. But transparent context names are still at risk of not being persistent.

Considerations for Managing Contexts
http://resolver.net.au/hdl/102.100.272/N8R5K6DQH

How should I manage multiple contexts?

Community Guidelines and Considerations

- Have a model of abstract contexts. Abstract contexts are defined by policies and owners, not identifier systems.
- Model hierarchies of abstract contexts relevant to your enterprise: one context can be a subcontext of another.
- Concrete contexts (real identifier systems) realise abstract contexts. Only deploy the system when you know what it is realising.
- A concrete context should be complete: it should realise all the identifiers in an abstract context.
- An abstract context can be realised redundantly in two systems (which means two actionable identifiers for the same thing). This increases

reliability and user base, but can also cause confusion and prevent metadata linking.

- If two concrete contexts are being used, one should be treated as preferred. The preferred context must be persistent.

Considerations for Managing Contexts
http://resolver.net.au/hdl/102.100.272/N8R5K6DQH

What authority metadata should be associated with any identifier created?

How do I keep identifiers accountable?

Community Guidelines and Considerations

- Authority metadata establishes identifier provenance and trustworthiness.
- Any identifiers used in authority metadata should themselves be persistent, or else non-actionable.
- The current manager of an identifier should be cited with a loose rather than a rigid identifier. (e.g. email alias, not personal email)
- Log identifier actions

Persistence of Identifiers Guidelines

http://resolver.net.au/hdl/102.100.272/V89DC0DQH

PILIN Policy – what we used in the PILIN Project

The authority metadata for an identifier shall consist of:

- *Date of creation of identifier*
- *Date of last update of identifier*
- *Identifier creator*
- *Identifier owner(s)* (*Description of content)*

Authority Metadata for the PILIN Project
http://resolver.net.au/hdl/102.100.272/D7J1G0DQH

Can I keep my existing identifiers if I move them to a new identifier scheme?

Community Guidelines and Considerations

Yes, as long as the new identifier scheme policy allows it. This is likelier if you have exclusive access to the new identifier system.

Persistence of Identifiers Guidelines

http://resolver.net.au/hdl/102.100.272/V89DC0DQH

Do I give identifiers to different versions of an object?

Do I give identifiers to different presentation?

Do I give identifiers to different copies?

Community Guidelines and Considerations

Only if the distinctions between them are meaningful to your business process, and they are likely to be cited as discrete things by external users.

(See also What to Identify?)

Identifier Association Policy for the PILIN Project
http://resolver.net.au/hdl/102.100.272/3SYFCJ3QH

How in Handles do I relocate my naming authority to a different institution?

Community Guidelines and Considerations

A handle needs to change its admin to a new namespace.

Possible solution: Use JAHDL/similiar technology.

Persistence of Identifiers Guidelines

http://resolver.net.au/hdl/102.100.272/V89DC0DQH

Do I have a single resolution service for my identifiers?

Community Guidelines and Considerations

- A single canonical resolution service is the default assumption for identifiers, inherited from URL.
- Multiple resolution services allow more leverage out of an identifier, and more flexibility.
- If there are multiple resolutions, one should be presented as preferred or default.

Identifier Services Guidelines
http://resolver.net.au/hdl/102.100.272/1KKBLPDQH

How do I make sure an Identifier is persistent while I have control over it?

Community Guidelines and Considerations

- Avoid patching (changing the identifier)
- Consider aliasing
- Avoid meaningful labels
- Map between multiple contexts
- Have an information model
- Do not reuse labels
- Use typed identifiers
- Use loose identifiers
- Validate identifiers
- Use redundant association data
- Embed identifiers into things

Persistence of Identifiers Guidelines

http://resolver.net.au/hdl/102.100.272/V89DC0DQH

How do I make sure an identifier is persistent if I no longer have control over it?

Community Guidelines and Considerations

- Delegate identifier management
- Federate identifier context
- Alias identifiers
- Patch identifiers (not recommended!)
- Make lifespan of context explicit
- Delegate Identifier context
- Alias identifier context
- Patch identifier context (not recommended!)
- Preserve provenance
- Mirror during transition of identifier
- Advertise move of identifier

Persistence of Identifiers Guidelines

http://resolver.net.au/hdl/102.100.272/V89DC0DQH

How do I get my own Handles?

Community Guidelines and Considerations

You will need to be allocated your own Handle namespace, which defines the Handle identifiers that you can manage yourself. The namespace is managed through your own Handle server.

http://handle.net/start.html

How do I ensure that a service operating on a persistent identifier (eg a resolver) is itself persistent?

Community Guidelines and Considerations

- Manage service separately from identifiers
- Avoid meaningful and branded names
- Have handover arrangements
- Host in large institution or consortium.

- Expose the fact that a service is preferred.

Identifier Service Guidelines
http://resolver.net.au/hdl/102.100.272/1KKBLPDQH

How can I use types or service identifiers with handles to ensure that handle identifiers is are actioned in a specific way?

Community Guidelines and Considerations

Use typed identifiers –a given identifier, once created, may only be associated with a particular type of thing.

Persistence of Identifiers Guidelines
http://resolver.net.au/hdl/102.100.272/V89DC0DQH

Shouldn't I just stick to URLs for my identifiers?

Community Guidelines and Considerations

- Whether the retrieval key for your resources is a URL, a Handle, or a folksonomy tag, you should still treat it an identifier, and manage it like one. The policy infrastructure needed to make Handles persistent is just as necessary to make URLs persistent.
- Users by default think of URLs as locations, not persistent identifiers. Dissociate persistent URLs from file names and from meaningful attributes in general.
- URLs are strongly bound with the DNS domain model, and with meaningful names for DNS domains. Meaningful context names work against persistence: the persistence of the URL is bound to the persistence of the domain name, and cannot be straightforwardly migrated to a different provider ("rehomed"),.
- Provide guarantee of trustworthiness: users expect URLs not to be persistent. Have barrier to entry for persistent URLs you provide.
- Avoid branded DNS domains.

Using URLs as Persistent Identifiers

http://resolver.net.au/hdl/102.100.272/DMGVQKNQH

Copyright © Monash University

This work is licensed under the Creative Commons Attribution-Share Alike 2.5 Australia License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/2.5/au/

This work was created as part of the PILIN project. The PILIN project is funded by the Australian Commonwealth Department of Education, Science and Training, (DEST) under the Systemic Infrastructure Initiative (SII) as part of the Commonwealth Government's Backing Australia's Ability – An Innovation Action Plan for the Future (BAA) under the ARROW Project.