

PILIN identifier for this work: PILIN/461BL3DQH

Handle for this work: hdl:102.100.272/461BL3DQH

 <p>pilin persistent identifier linking infrastructure</p>	<p>web: http://resolver.net.au/hdl/102.100.272/0N8J991QH email: policy@pilin.net.au</p>
---	--

Version History

Version	Date	Status & changes	Expression identifiers
V1.0	2007-12-17	Release	PILIN/ TF17KVNQH hdl:102.100.272/ TF17KVNQH

PILIN Project Guidelines

Considerations for Ownership of Identifier Management Systems

This document is a work in progress and may contain open questions not resolved during the timeline of the PILIN project. It represents the thinking of the PILIN team as at December 2007.

To cite the *latest* version of this work use <http://resolver.net.au/hdl/102.100.272/461BL3DQH>

To cite *this* version of this work, use <http://resolver.net.au/hdl/102.100.272/TF17KVNQH>


1 Purpose/Issue

This guideline document goes through considerations on what arrangements parties should make for owning and managing identifier management systems—whether they should manage the system themselves or jointly with other parties, and whether they should have exclusive access to the identifiers managed through the systems.

2 Background

An identifier management system, as defined in the PILIN information model, is a collection of definitions, information models, policies, and data sources to manage identifiers. Once data sources are used to manage identifiers, an access boundary is defined: only authorized parties have access to things inside the boundary. The things of interest in this case

Copyright © Monash University

 <p>CC BY SA</p>	<p>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/</p>
---	--

This work is created as part of the PILIN – Persistent Identifier Linking Infrastructure project. The PILIN project is sponsored by the Australian Commonwealth Department of Education, Science and Training under the ARROW Project.

are the identifiers being managed, and the access type of interest is write access to the data sources.

Identifier management systems are themselves owned by parties: those parties assume responsibility for the system and its ongoing operation. We refer to this type of ownership as *hosting*. Who hosts the system is independent of who can manage identifiers through the system.

If a party is authorized to manage identifiers through a system, there are two logical alternatives: either only one party is authorized to use the system (*exclusive management*), or else several parties share access to the system (*shared management*).

There are also two logical alternatives for hosting: either the party managing the identifiers also hosts the management system (*own hosting*), or else some other party does (*external hosting*). If the party owns the system jointly with others, we consider that to be own-hosting.

These two choices lead to four possible arrangements for hosting identifier management systems and managing identifiers, which have different implications for persistence and policy.

- *Exclusive management, Own hosting*: identifier management is **devolved**. Each naming authority has complete control over its identifiers, and there is no pooling of system resources.
 - E.g. the physics department of a university hosts its own identifier system, PHYS-ID, and the chemistry department of the university hosts its own identifier system, CHEM-ID. Physics staff cannot edit CHEM-ID identifiers, and Chemistry staff cannot edit PHYS-ID identifiers.
- *Shared management, External hosting*: identifier management is **centralised**. In the centralised case, a single naming authority hosts identifiers from several institutions in a common system; that is, the centralised naming authority provides the system, and the institutions manage their identifiers through the central system. The identifier context is pooled between managing parties.
 - E.g. the physics department and the chemistry department of the university both use identifiers hosted by a central, institutional system, UNI-ID. Both Physics and Chemistry staff can manage UNI-ID identifiers, and neither has an identifier context unique to their department.
 - Identifier management can also be **surrogate**: one institution hosts another institution's identifiers, by arrangement between the two. The distinction between centralised and surrogate does not affect the considerations discussed here.
- *Exclusive management, External hosting*: identifier management is **autonomous**. A centralised naming authority hosts identifiers

from several institutions, but each institution is provided with its own host. The identifier context is unique to the managing party.

- E.g. the physics department manages identifiers through one identifier system, PHYS-ID, and the chemistry department manages identifiers through a different identifier system, CHEM-ID. Each system is specific to that department. But both systems are hosted centrally by the university, not by the departments.
- *Shared management, Own hosting*: identifier management is **federated**. Multiple naming authorities exist, and each hosts their own server; but the naming authorities are in a trust relationship (e.g. a consortium), so they can all manage identifiers on the common system. The consortium provides its own hosting infrastructure.
- E.g. the physics, chemistry, and biology departments form an identifier consortium. Each department has its own identifier system. But because of the common trust environment, a physics staff member can manage CHEM-ID identifiers by arrangement. PHYS-ID and CHEM-ID are not hosted separately: they are hosted jointly by the consortium members.

3 Scope

These guidelines inform the choice of identifier management and hosting model according to the two criteria described: who hosts the identifier system, and who else is allowed to manage identifiers through the system. They are not specific to any identifier technology. They do assume that the identifier management systems are closed and not publicly writeable. The considerations given are not exhaustive, and concentrate on policy control, persistence, and branding issues.

4 Guidelines: Own Hosting

Do I host my own identifier management system, or use someone else's?

YES	NO
I have my own identifier namespace and management system, hosted on my own server onsite	I use someone else's identifier management system. But I still have access to edit my own identifiers.

If I host my own:

Pros	Cons
I have maximal control over my identifiers, and the services I wish to associate with them	I shoulder the administrative and technical burden of running an identifier management system, and all that entails (reliability, backup, mirroring, etc)
Minimal technical or administrative overhead on external identifier authorities	I have sole responsibility for the Handle server: failure can happen much more easily
I can still benefit from external identifier expertise	I may not benefit from the standard interfaces, name interpretations, and procedures of a well-established identifier system
I can customise identifier solutions to my requirements	

In summary: hosting my own identifier system gives me full control over what I can do to and with my identifiers. But it also commits me to the full workload of running such a system, and can also move me away from standard identifier solutions.

5 Guidelines: Exclusive Management

Am I the only party with access to the management system, or am I sharing it with someone else?

YES	NO
I have exclusive access to my identifier namespace (i.e. the contents of the identifier records being hosted), and no one else can add or modify identifiers in that namespace	I am sharing the identifier namespace with some other party

If I have exclusive management access to my system:

Pros	Cons
I can set up a clear governance model for my identifiers, without having to accommodate other parties' requirements	I cannot manage transition of someone else's identifiers, in case they move into my identifier system —I manage the new system but not the old

I can coordinate my identifier workflows with my other workflows, without having to accommodate other parties' requirements	As a result, party-specific identifiers may not persist if an item migrates between parties
I can manage transition of my own identifiers closely, in case I move them to a new identifier system—I manage both the old and the new system	I constitute a single point of failure for identifier management: no other parties can share the load
I can devise my own label policy without needing to prevent collision from other parties' labels	
I can have strong branding of my identifier context: the context really is mine alone.	

In summary: if I have exclusive control of the identifiers in my system, then I do not have to negotiate with other parties on how to set up my identifiers, I do not have to “dumb down” my identifiers to a lowest common denominator, and I can brand my identifiers as truly my own. These are compelling arguments, and indeed identifiers are usually managed exclusively.

There are two reasons why management might be shared. The first, as for hosting, is to provide added insurance of persistence: if several parties have access to manage identifiers, then one can take over management from another with minimal disruption.

The second, related reason is to address the use case of persistent identifiers when the thing identified passes out of the control of one party and into another (see “Persistence of Identifiers Guidelines: Association with Things outside one’s Control”: “Guidelines: Thing Moves Away”). For example, an academic moves from Melbourne Uni to Monash Uni, takes their papers with them, but wishes the Melbourne-assigned identifiers for those papers to remain persistent. Normally this can force a new identifier to be created, as the old identifier can no longer be maintained by the original party. But if identifier management is shared between the old and the new manager of the thing, then both parties can update the identifier, and there is no need to create a new identifier.

6 Management profiles

We consider more specific pros and cons for each of the four management profiles identified above.

- The default profile is devolved management, where each party manages their own identifiers and identifier systems.
- Centralised systems introduce an economy of scale, and relieve parties of administrative burdens; but they also take away much of the ownership of the identifiers being pooled into the centralised system.
- Autonomous systems address this problem by uncoupling hosting from identifier management.
- Federation addresses the problem in a different way, restoring ownership through a shared consortium—but this introduces its own administrative burden at the consortium level.

6.1 Devolved management

- Each party hosts its own system, and has exclusive access to it.
- Any central identifier authorities have at most only an advisory/coordinating role. However they may host value-added identifier services

Pros	Cons
Individual parties manage and host their own identifiers, without surrendering any control	Individual parties are more vulnerable to failure than under federated or centralised profiles
Minimal technical or administrative overhead on any external identifier authorities	Failure of identifier management by individual party reflects badly on the party—and also on the chosen identifier infrastructure (especially if it has been marketed as persistent)
Parties can still benefit from external identifier authority expertise	Migrating identifiers from a different party and keeping them persistent is problematic—requiring a centralised or federated service mapping between identifiers
Parties can customise identifier solutions to their requirements	
Parties have complete control over their label policy	
Parties can easily persist old identifiers they manage	

Strong branding of identifier context (namespace) is possible	
---	--

6.2 Centralised (Surrogate) management

- Identifiers for various parties are hosted together within a single system, on a single host.
- Identifiers are typically hosted by a dedicated third party.
- Identifiers for various parties are not managed discretely.

Pros	Cons
Individual parties need not provide their own hosting infrastructure	Individual parties lose control over their own identifier management. Identifiers must comply with central policy, any changes to the identifier must be mediated through the central host, and any changes to the infrastructure must be approved by the host
Things can move from one party to another without extensive change needed in identifier management, as long as the parties both use the central host. As a result, identifiers can persist through migration between those parties, or at least a centralised mapping service between identifiers can be provided	No branding or explicit clustering is possible for identifiers as belonging to a particular party
	Significant administrative overhead on central authority, including negotiating mutually acceptable policies and infrastructure
	Other parties' labels in the same context may give rise to collisions; avoiding this requires additional infrastructure
	Any identifiers the party already manages as specific to them need to be transitioned to the new centralised system—breaking their persistence

6.3 Autonomous management

- Each institution has its own exclusive namespace.
- The systems managing the namespaces are hosted centrally

Pros	Cons
Individual parties manage their own identifiers, without surrendering any control over the identifiers	Individual parties lose control over their own identifier <i>system</i> management. Any changes to the identifier infrastructure (e.g. services, SLAs) must be approved by the host
Individual parties still manage their own identifiers separately from other parties	Significant administrative overhead on central authority, including negotiating mutually acceptable infrastructure, and maintaining multiple distinct identifier systems
Individual parties need not provide their own hosting infrastructure	Weakened branding of identifier contexts (namespaces): the individual identifiers are the party's own, but the identifier host is not
Things can move from one party to another without extensive change needed in identifier management, as long as the parties both use the central host. As a result, identifiers can persist through migration between those parties	Migrating identifiers from a different party and keeping them persistent is problematic—requiring a centralised or federated service mapping between identifiers
Parties have complete control over their label policy	
Parties can easily persist old identifiers they manage	

6.4 Federated management

- Parties contribute their own hosting infrastructure, and have their own namespaces
- Hosting is undertaken as part of a federation
- Parties share namespace management

Pros	Cons
Individual parties manage and host	Any changes to the identifier or the

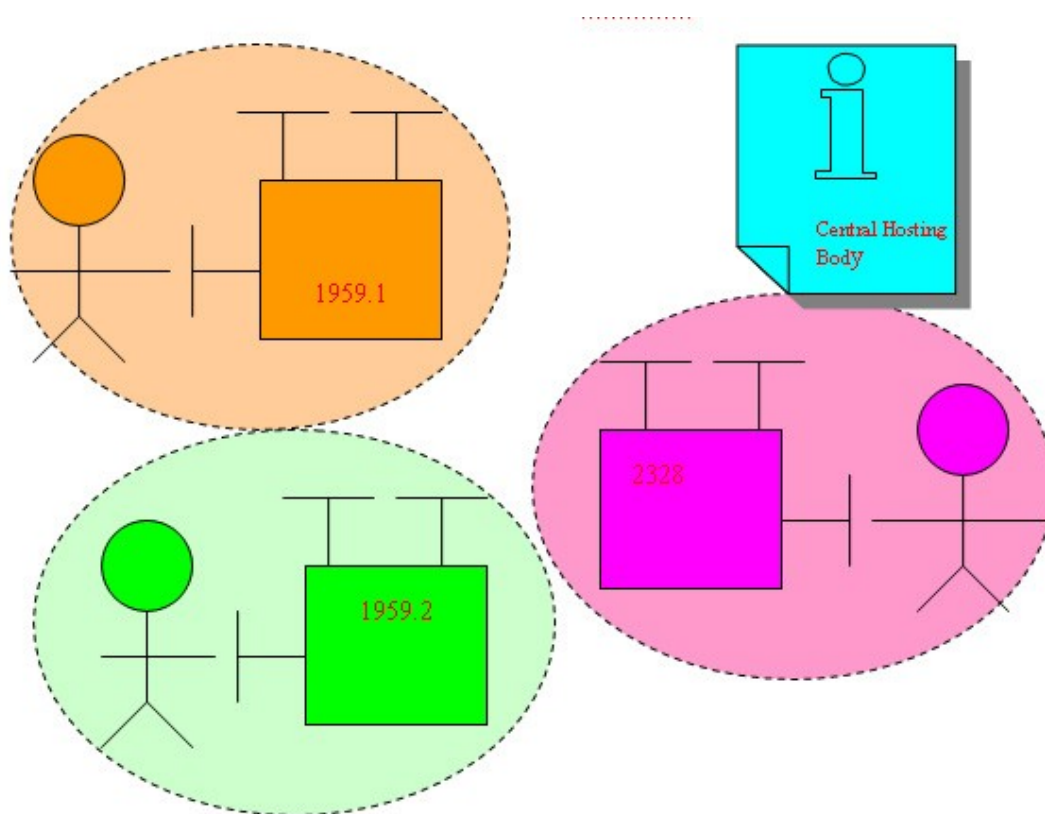
their own identifiers	identifier infrastructure must be negotiated within the federation
Minimal technical or administrative overhead on any external identifier authorities	Individual institutions do not have exclusive access to their identifiers (they do not "own" them). As identifiers migrate between parties over time, the ownership of a particular identifier becomes unclear. Authority over identifiers (including accountability) is generally decentralised and harder to track
Parties can still benefit from external identifier authority expertise	Failure of identifier management by a party within the federation reflects badly on the entire federation (since the federation has assumed corporate responsibility over the identifiers)
Parties can customise identifier solutions to their requirements	It remains problematic to maintain persistent identifiers if they refer to objects moving outside the federation
Parties can easily persist old identifiers they manage	Shared trust infrastructure needs to be deployed (federation introduces extra complexity of system). The infrastructure needs to be institutionally guaranteed over long timespan
Strong branding of identifier contexts	Difficult to leave a trust federation once entered: individual ownership of identifiers has been compromised
Parties have complete control over their label policy	
Individual parties work within a consortium, can rely on each other for long term persistence	
Things can be migrated to administratively distinct parties within the federation without any change necessary in the management of the identifier. As a result, identifiers can persist through migration between those	

<p>parties, or at least a centralised mapping service between identifiers can be provided</p>	
---	--

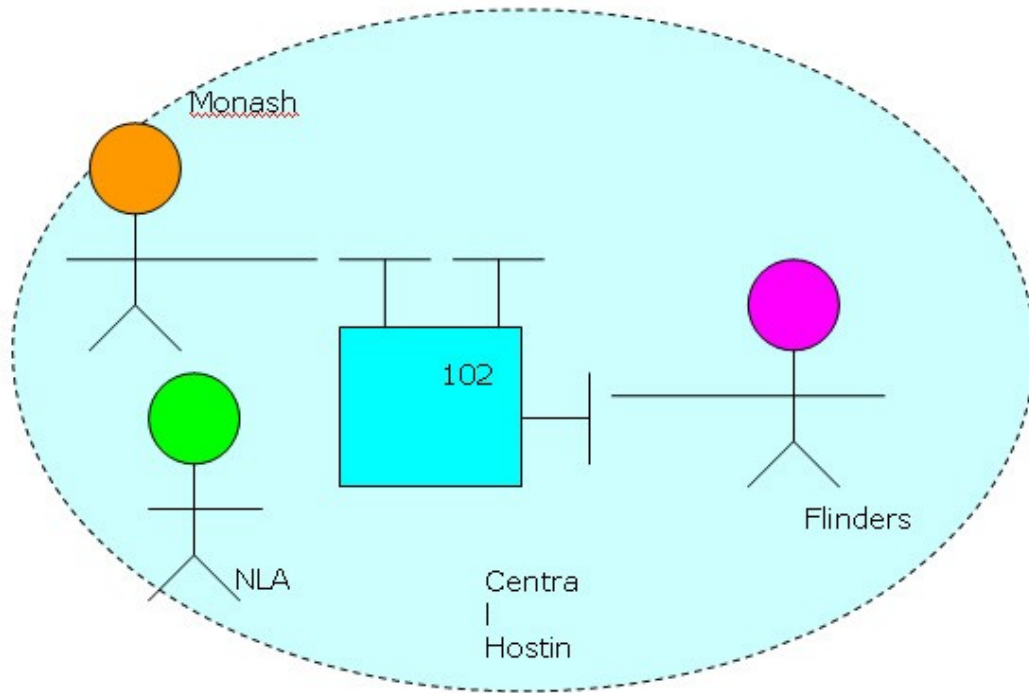
7 Example

Diagrammatical representation of each hosting scheme:

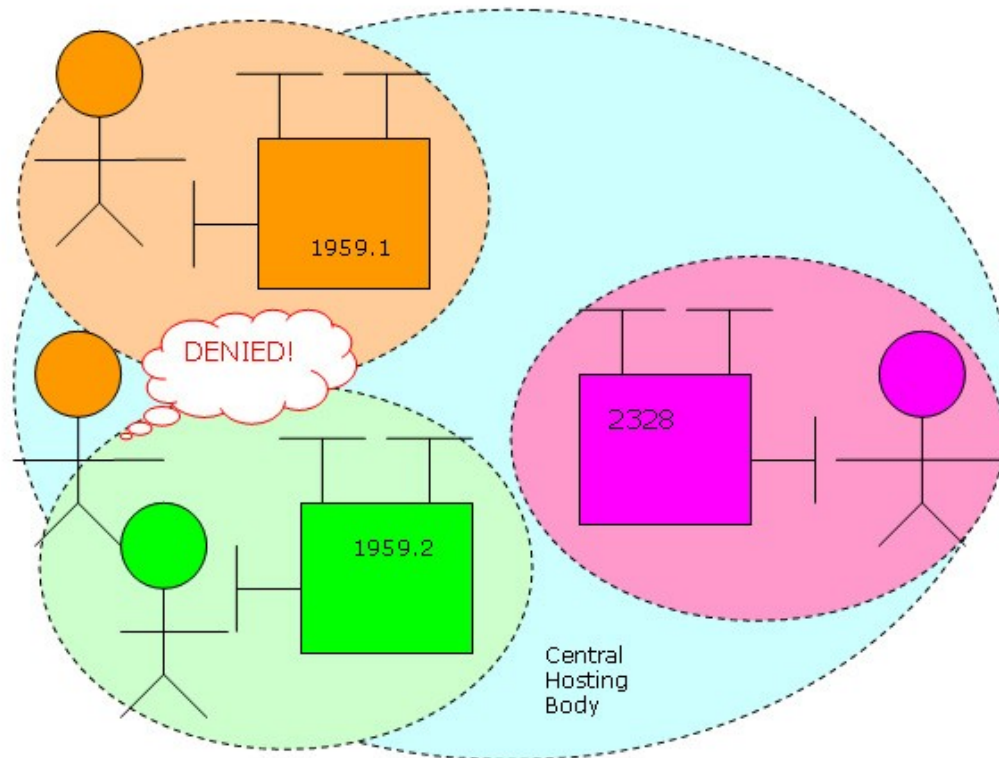
7.1 Devolved: own hosting, own curating



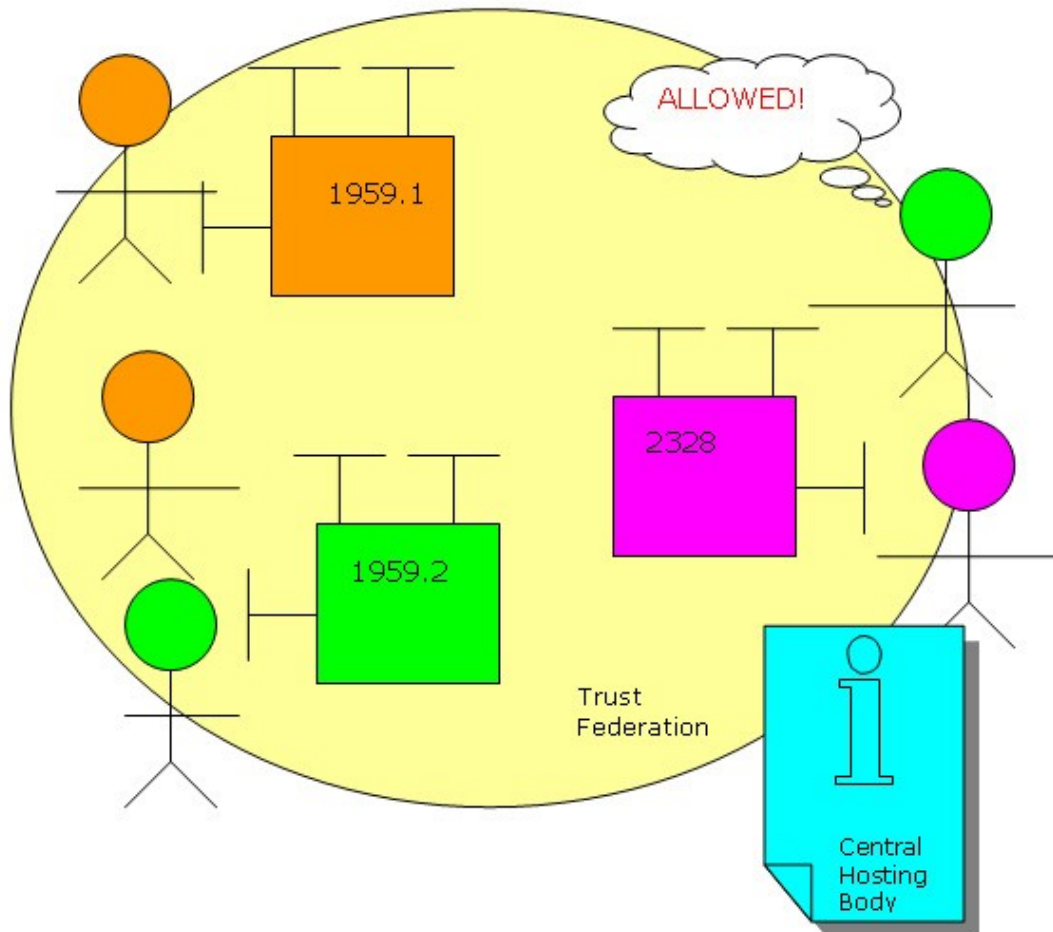
7.2 Centralised: external hosting, shared curating



7.3 Autonomous: external hosting, own curating



7.4 Federated: own hosting, shared curating



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.

