Persistent Identifier Linking Infrastructure (PILIN)

Stakeholder Meeting
March 2007

The ARROW Project is funded by the Australian Commonwealth Department of Education, Science and Training, under the Research Information Infrastructure Framework for Australian Higher Education.

arrow.edu.au

The ARROW Consortium comprises Monash University [lead institution], National Library of Australia, The University of New South Wales and Swinburne University of Technology.

WELCOME

PILIN objectives

Pilot shared, standards-based, persistent identifier management infrastructure

Support adoption of

- persistent identifiers
- persistent identifier management services
- by Australian e-learning, e-research and e-science and cultural communities
- through "proof of concept" deployment and demonstrator services

Plan for sustainable shared infrastructure

- persistence of identifiers
- persistence of identifier services
- over archival lengths of time

Why PILIN?

- Managing and discovering digital (and other) resources as they move around during their lifecycle
- Establish governance and policy
  - as important as technology in guaranteeing persistence
- Response to growing realisation of importance of global identifiers
- Provide shared identifier infrastructure for a range of projects
  - avoiding duplication and enabling interoperability
  - Basic infrastructure for repository projects

Why shared management infrastructure?

Return on investment

- Co-develop and reuse tools and services
- Interoperate with other information infrastructure

Participating across a global networked environment

- e.g. global access to rapidly migrating and proliferating resources

Assurance

- Shared responsibility for uniqueness of identifiers and persistence of identifiers and identifier services
- NB: requires pro-active, trusted ownership and shared responsibility

Scalability to millions (billions?) of identifiers

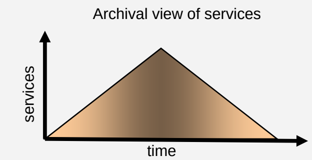
What is identifier management infrastructure?

Information infrastructure and services built using that infrastructure

- Core management identifier services
  - Create, access, update
- "Value-added" resource management services
  - Create version, Find appropriate copy, Move resource, …

The services:

- are dynamic
  - e.g. Find current location of resource
- require identifiers that are:
  - globally unique
  - persistent over time
- should themselves provide reliable action over time



Why based on CNRI Handles System?

- Handles in use now
  - by MERRI and FRODO Projects
  - Grid integration underway
- Extensible and scalable architecture
  - we can build other (web) services on top of the core services
  - demonstrated ability to build a scale
- Deals with non-web resources
- Allows Australia to be both part of global infrastructure and be independent
- Handle is one instantiation of an abstract identifier
  - Want to integrate other identifier services

Get abstractions, practices and governance "right"

- expect technology will migrate over time

Story so far

- Assembled the team
- Started consultation
- Initiated overall approach
- Documented initial thoughts and ideas
- Developed a development space to work in

Stakeholder engagement

- **Today is not the beginning**
- **Today is not the end**
- **Today is not the totality**

http://www.arrow.edu.au/PILIN.php

Project outputs

1. **Model** lifecycle of persistent identifiers and services
2. Best practice and policy **guides**
3. Document community **requirements**
4. e-Framework representations of identifier management **services**
   - Technology-independent *service genres*
   - Technology-dependent *service expressions*
5. **Pilot** shared persistent identifier management infrastructure
   - Using Handles system
6. **Software toolkits** to ease use of the infrastructure
7. **Sustainability** options and proposals

http://www.arrow.edu.au/PILIN.php

The Pilot Infrastructure

Develop using *agile* software development methodology

Priorities:

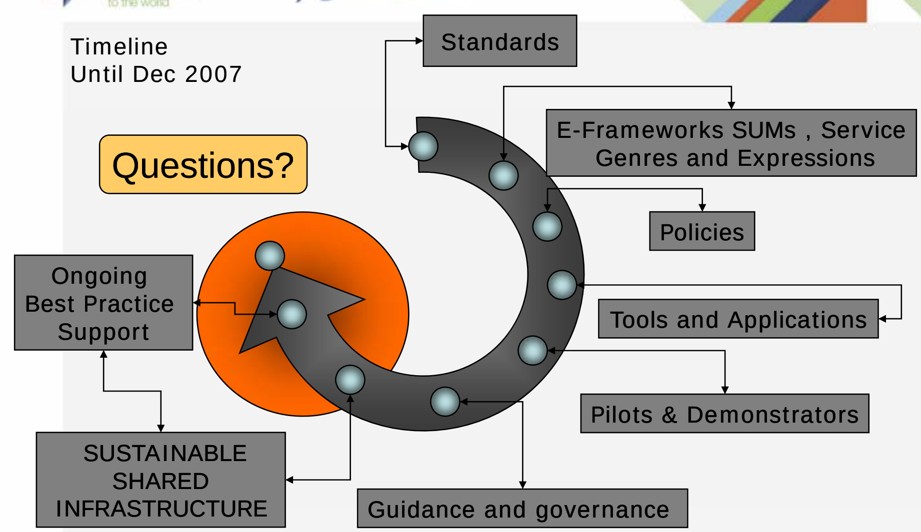
1. Base infrastructure: CNRI Handles System
2. Core management services
3. Toolkits
4. Interoperation with other systemic information infrastructure
   - AAA, GRID Handles, NLA identifiers
5. "Value-added" resource management services

Access

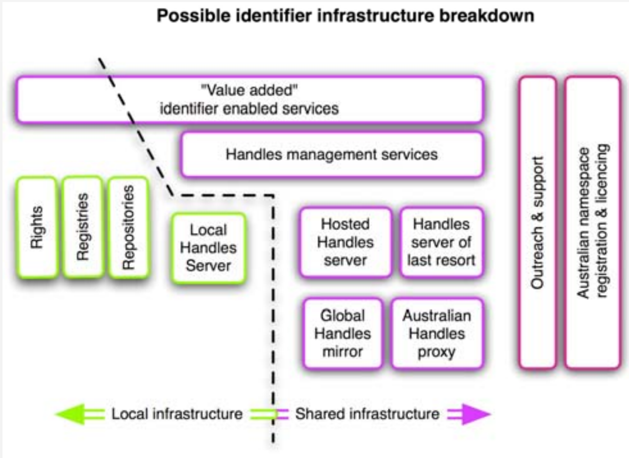
- For exclusive use of project stakeholders
- Piloted by selected stakeholders
  - FRODO and MERRI projects (set of DEST funded projects)
  - Large e-science data collection
  - Schools, VTE and transport sectors

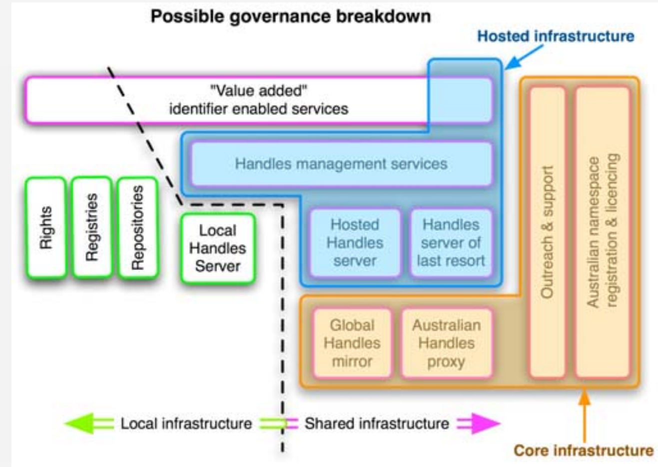
http://www.arrow.edu.au/PILIN.php

Timeline
Until Dec 2007



http://www.arrow.edu.au/PILIN.php





Framework for identifier qualities

Define some commonly used terminology:

- Identifier
- Identifier system
- Registered
- Accountable
- Resolvable
- Actionable

It is hard!

- Some early thoughts …

What is an identifier?
Our working definition

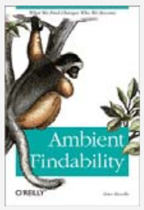
A label	0DTRD1Q3C
In a unique context	PILIN project identifier system: 0DTRD1Q3C
Associated with an object	http://www.pilin.edu.au/documents/best-practices.html

What is an identifier?
Our working definition

A label 0596007655

In a unique context 10 digit ISBN: 0596007655

Associated with an object Ambient Findability, Peter Morville



What is an identifier?
Our working definition

A label 414 555 555

In a unique context Phone number: +61 414 555 555

Associated with an object SIM card

Realisation of identifiers

Identifier systems

- Manage labels
  - Strings, suffixes, …
- Manage contexts
  - Namespaces, domains, …
- Manage object associations

For example

- PILIN project identifier system: 0DTRD1Q3C
- Handles system with namespace 102.100.272
  hdl:102.100.272/0DTRD1Q3C
- PURL system with domain purl.pilin.edu.au
  http://purl.pilin.edu.au/0DTRD1Q3C

Identifier qualities

Registered

- Registered with an *identifier system* that manages the namespace

Accountable

- The *identifier system* can provide information on
  - how the identifier came to be
  - how it can be managed
- Requires
  - Policy
  - Policing of policy by the *identifier system*

Identifier qualities

Actionable

- There is a *service* that takes an identifier and performs an action

```
Appropriate_copy(102.100.272/0DTRD1Q3C) -> closest
  copy of the object

Billing(414 555 555) -> latest phone bill
```

- Often confused with web resolvable / clickable …

Identifier qualities

Resolvable

- Subset of actionable …
- There is a *service* that takes the identifier and returns information on how to gain access to the associated object

```
Handles_resolve(102.100.272/0DTRD1Q3C) -> URL

Locate(414 555 555) -> SIM card + cell location
```

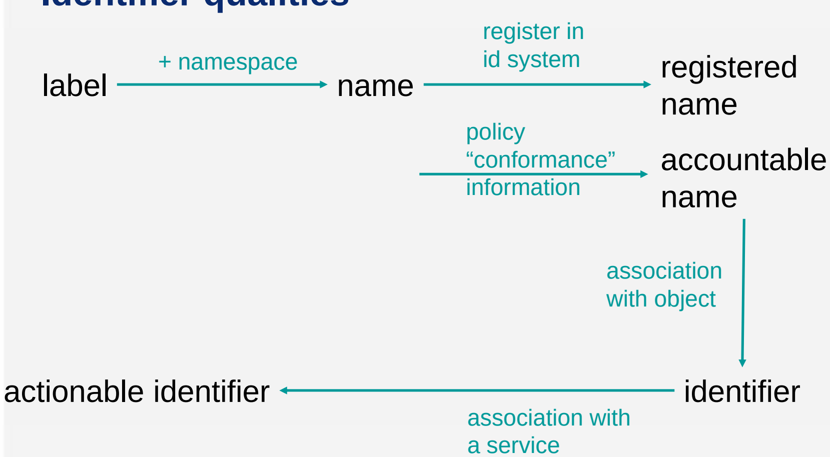
Web resolvable

= resolvable + URI encodable + returns a URL

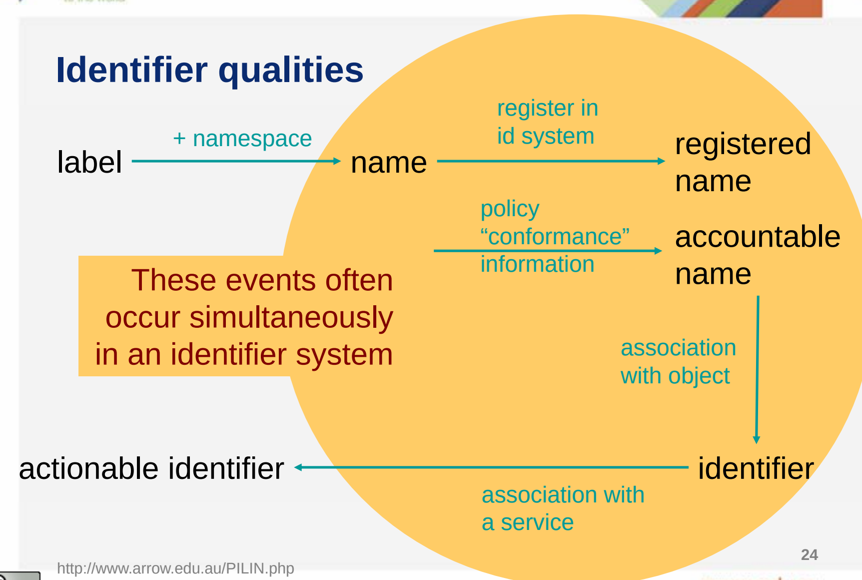
- “clickable” in a web page

```
http://arrowdev.lib.monash.edu.au:8081/
  102.100.272/0DTRD1Q3C
```


Identifier qualities



Identifier qualities



Identifier qualities

Internet encodable

- Encodable in a internet identifier system
  - Usually as UTF-8 octets
  - 102.100.pilin/0DTRD1Q3C

Web encodable

- Encodable as a URI
  - hdl:/102.100.pilin/0DTRD1Q3C
  - info-uri:hdl//102.100.pilin/0DTRD1Q3C
  - http://handle.net/102.100.pilin/0DTRD1Q3C
- Are these the same identifier?
  - Depends on your information model!

Stakeholder engagement

- **Today is not the beginning**
- **Today is not the end**
- **Today is not the totality**

Thank You!