



australian research repositories online to the world

Persistent Identifier and Linking Infrastructure (PILIN)

within the ARROW2 project
In Partnership with USA

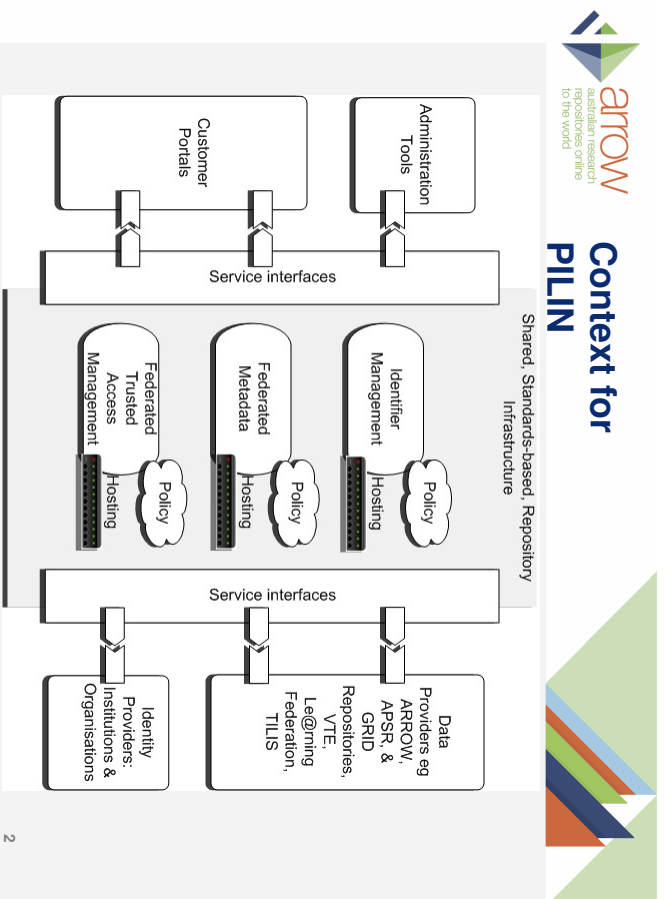

Dennis Macnamara
January 23 2007

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, Queensland National Library of Australia, The University of New South Wales and Swinburne University of Technology.





australian research repositories online to the world

E-Framework for Education and Research and PILIN

An initiative of the Australian Government Department of Education Science and Training and the UK Joint Information Systems Committee

Primary goal of the e-Framework is to facilitate technical interoperability within and across education and research through improved strategic planning and implementation processes.

Projects are closely linked



australian research repositories online to the world

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**

Plan for sustainable shared infrastructure

- **ersistence of identifiers**
- **ersistence of identifier services**

over **archival lengths of time**

Why PILLIN?

- Finding digital resources as they move around during their lifecycle
- Complements Repository projects
- Good governance and policy as important as technology
- Leveraging individual projects and economies of scale/return on investment
- Timely

5

Why shared management infrastructure?

- Return on investment
 - Co-develop tools and services
 - Interoperate with other systemic 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

6

What does this experience tell us? (1)

- Identifiers are needed for
- more than resolving to a web location
 - more than just web resources

- Need to go back to basics
- what are the characteristics of “identifier”?
 - what do we mean by “identifier”, “persistent identifier” etc
 - what are is the difference between an identifier and data associated with an identifier?

7

What does this experience tell us? (2)

- Need to decouple some information management concerns
- resource identification from resource location
 - associative data about identifiers from metadata about the resource
 - identifiers as abstractions from the technologies that instantiate them
 - information models and architectures that take advantage of identifiers from the identifiers

8

What does this experience tell us? (3)

- Persistence is as much about
- **governance**
 - **persistence of the basic management and information services** as it is about technology.

Need to provide Guidance and Good Practice to projects

Need good tools to help manage identifiers over their lifecycle

Great potential to provide interesting information services over identifier enabled infrastructure

- **enable identifier service mash-ups?**

9



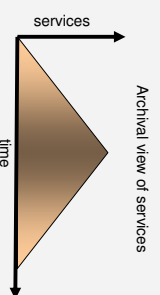
What is identifier management infrastructure?

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**



11



Why Handle based?

Handles are used now by other projects across educational and other communities eg NLA

CNRI Handle System is an architecture, not just a registration and resolution service

- **we can build other services on top of the core services**

Deals with non web resources

Proven scalability

Demonstrated ability to build a scalable Web Service layer on top

Architecture allows Australia to be both part of global infrastructure and be independent

Handle is one instantiation of the abstract Identifier

- **Want to integrate other required identifier services**

Get abstractions, practices and governance “right” - expect technology will migrate over time - 10

10



Australian experience with identifiers

FRODO and MERRI repositories:

- **Many repository projects use persistent digital identifiers to manage large numbers of networked resources**
 - e-science, e-research, and e-learning

- **Need for guidance and tools**

Emerging requirements

- **Vocational Training, Schools, Transport Industry**

Wide range of identifiable “resources”:

- **web content, data sets, scientific instruments, services ...**
- **emerging uses eg semantic web abstractions, Parties National Library**

- **persistent identifiers + information model that expresses relationships using identifiers = interesting services**

12



Project outputs

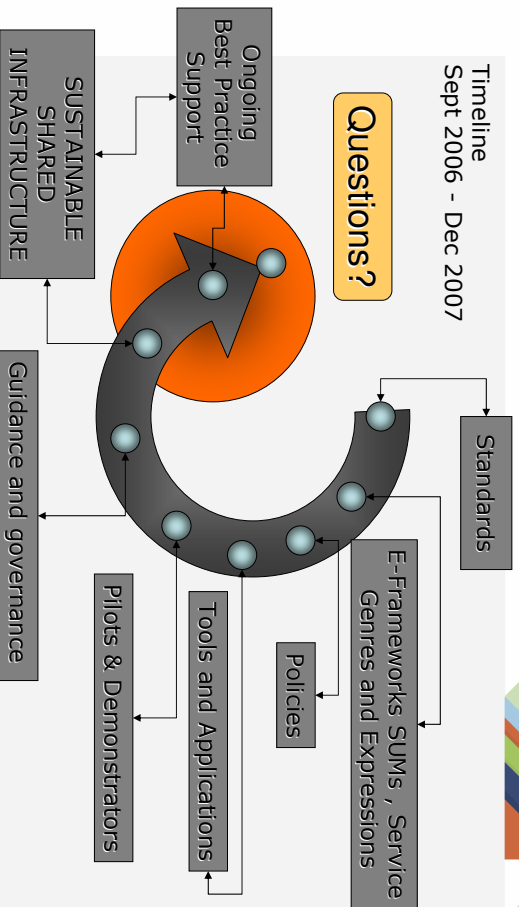
- Best practice and policy guides
- Document community requirements
- e-Framework representations of identifier management services
 - **Technology-independent service genres & technology-dependent service expressions**
- Pilot shared persistent identifier management infrastructure
- Toolkits to ease software use of the infrastructure
- Options and proposals for sustainability

13



PILIN – Summary

Timeline
Sept 2006 - Dec 2007



15



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 I/MERL projects (set of DEST funded projects)
 - Large e-science data collection
 - Schools, VTE and transport sectors

14



Some Issues

- Need to engage beyond education but limited resources in 2007
- Need to identify funding for real development in 2008 and beyond that meets the diverse needs of all education sectors, cultural institutions and other government and creative agencies
- Other big issue to get to grips with is IP management, but that is another presentation

16



Summary

- Adoption of Handle System based identifiers and identifier services
- Persistence and Sustainability plans
- Applicability to all educational sectors and to cultural institutions

17



SOFTWARE RESEARCH

arrow.edu.au

Follow up/more information

<http://www.arrow.edu.au/PILIN>

Dennis Macnamara

0412133175

ul@people.net.au

18



SOFTWARE RESEARCH

arrow.edu.au