

Persistent Identifier and Linking Infrastructure (PILIN)*

*PILIN is ARROW in Old Elvish

within the ARROW2 project

Nigel Ward, Technical Director
10 October 2006

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.



arrow

australian research
repositories online
to the world



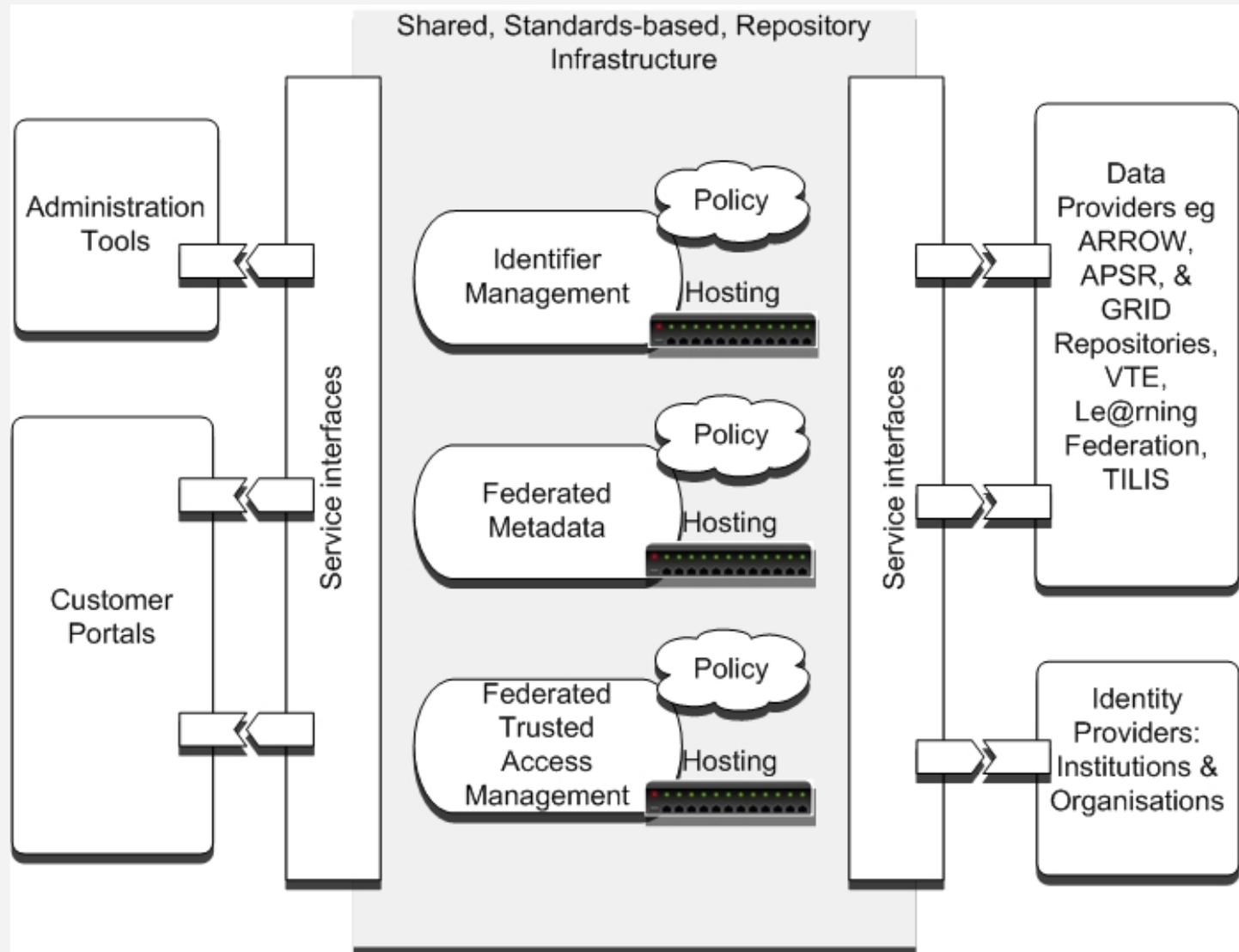
MONASH
University



UNSW



DEST context for PILIN



PILIN objectives

Pilot shared, standards-based, persistent identifier management infrastructure with the aim of

Supporting adoption of

- **persistent identifiers**
- **persistent identifier management services**
- **for Australian e-learning, e-research and e-science communities**

Planning for sustainable shared infrastructure

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

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**

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?**

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**

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!?**

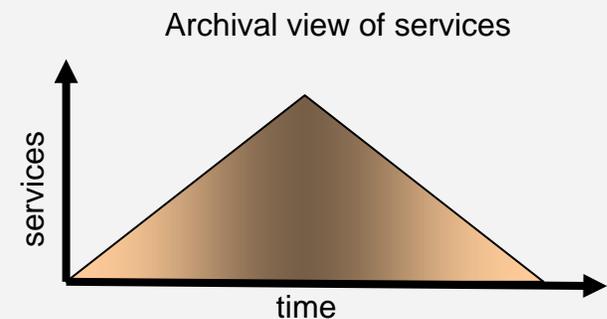
What is identifier management infrastructure?

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 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**

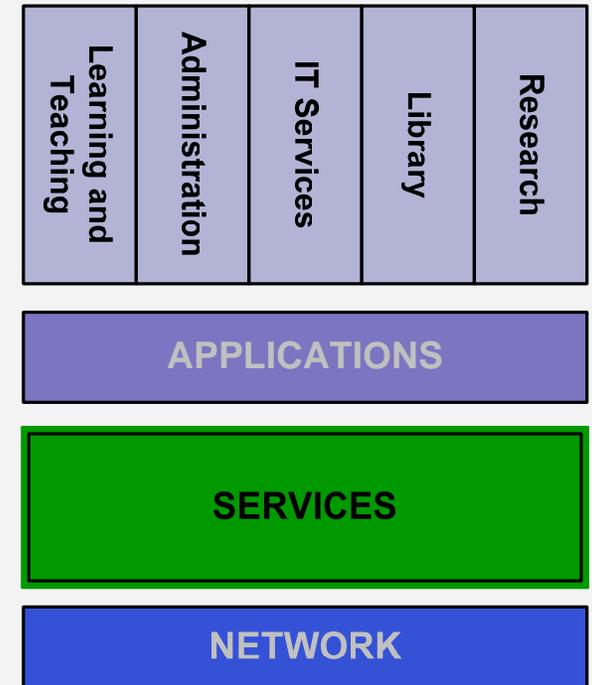
Scalability to millions (billions?) of identifiers

E-Framework approach

Open standards based interfaces
for interoperability
In a service oriented environment

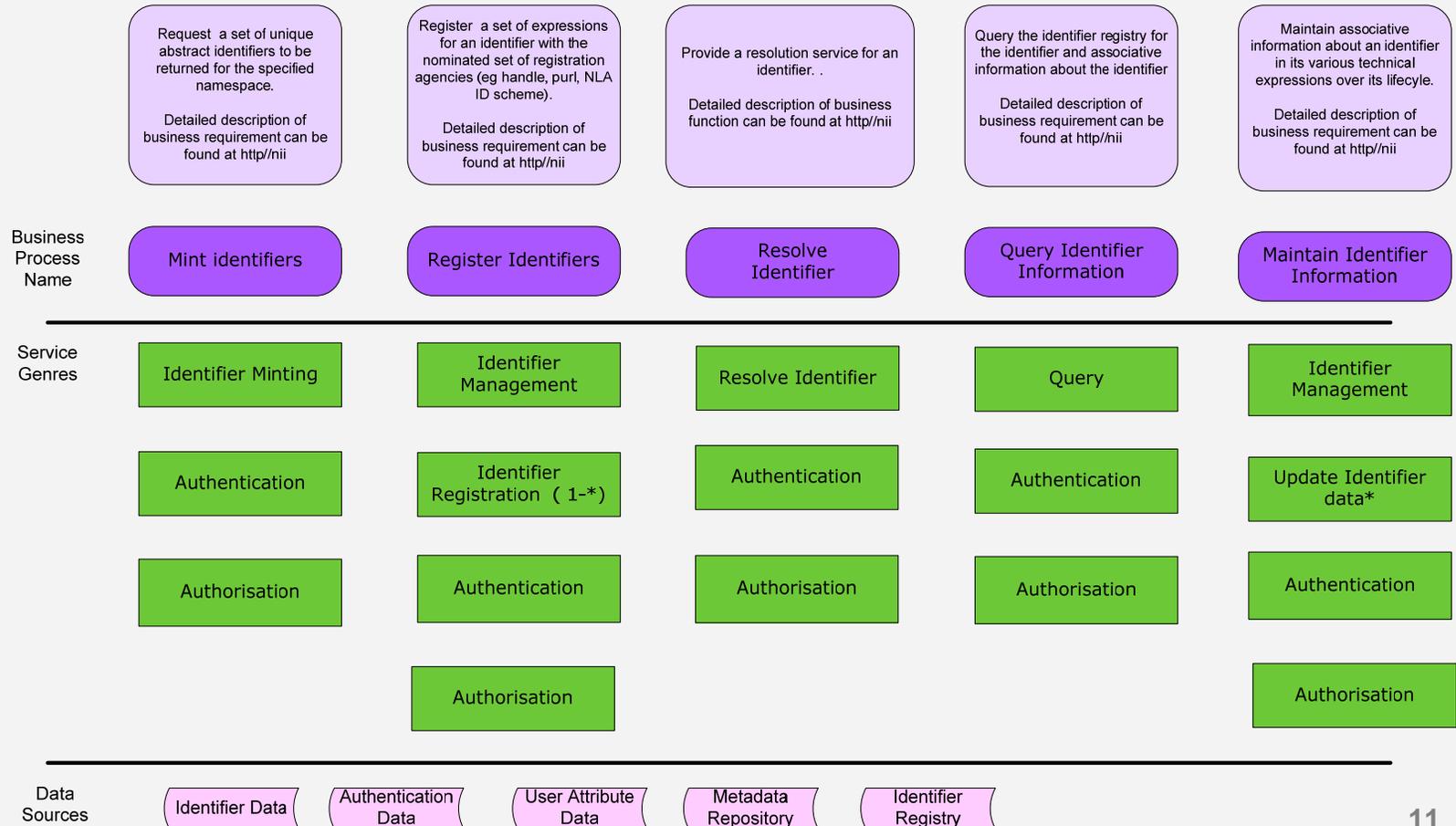
The e-Framework analyses and
documents *SERVICE INTERFACES*

*PILIN WILL DEVELOP SERVICE
INTERFACE DEFINITIONS and
TOOLKITS to support IDENTIFIER
MANAGEMENT*



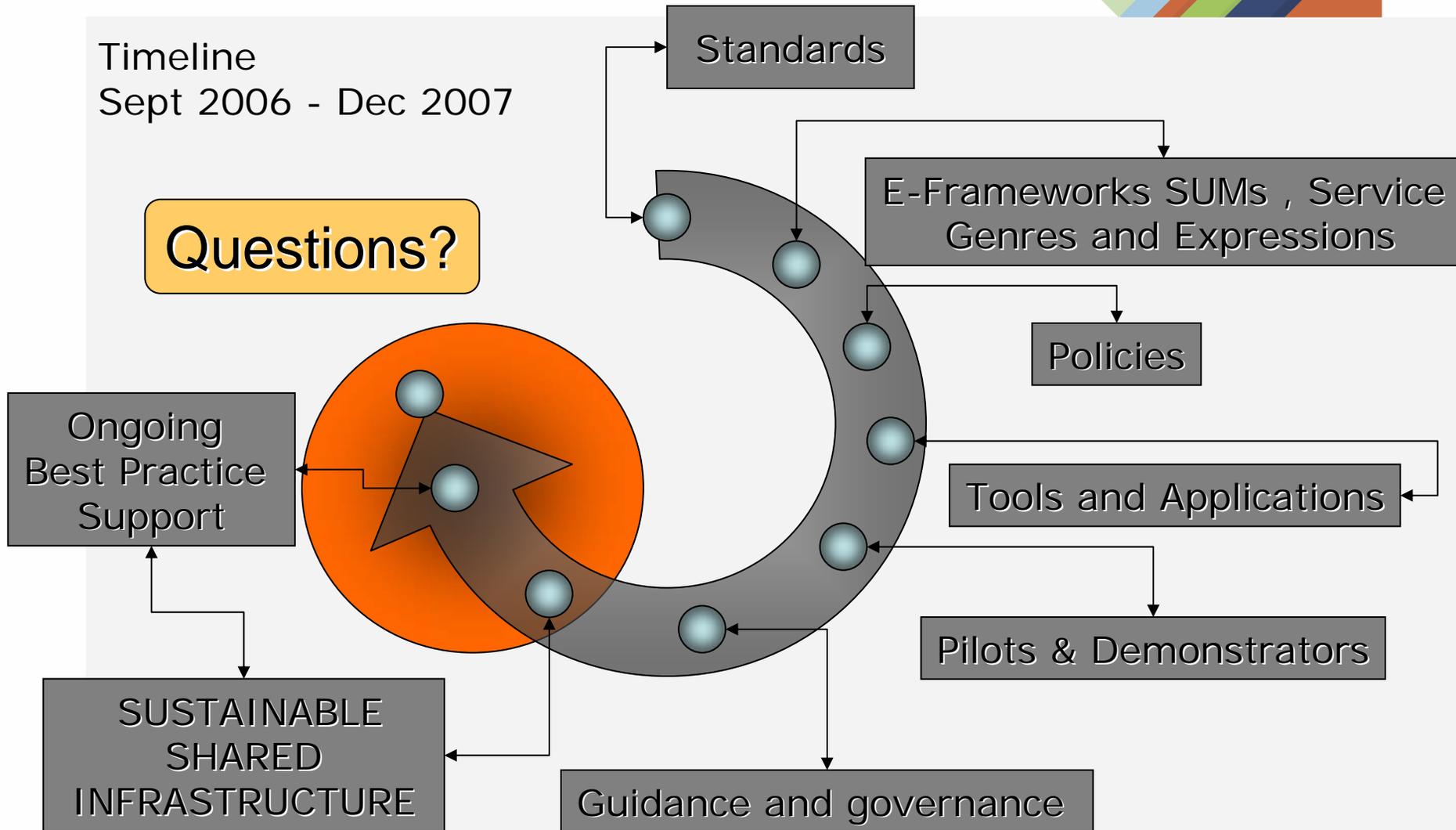
Identifier Management Example – note we know this isn't right!!!

IDENTIFIER MANAGEMENT Service Usage Model



Timeline
Sept 2006 - Dec 2007

Questions?



Why Handle based?

Handles are used now by the MERRI and FRODO Projects

Grid integration underway

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.