PILIN Project: Project Closure Report

PILIN Team

Date: 20/12/2007

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

Table of Contents

1 Overview of Project..... 1
1.1 Background..... 1
1.1.1 Why identifiers? 1
1.1.2 Why identifier infrastructure? 2
1.1.3 Why shared identifier management infrastructure? 2
1.2 Aims and Objectives..... 3
1.3 Project Governance..... 3

2 Project Outputs..... 6
2.1 Abstract Model for Persistent Identifiers – Ontology..... 6
2.2 Community Requirements..... 6
2.3 Advice on Using Identifiers..... 6
2.3.1 Internal Policy Documents 6
2.3.2 Guidelines and Considerations.....6
2.3.3 PILIN Glossary.....7
2.3.4 Identifier Policy FAQ.....7
2.4 Service Oriented Models..... 7
2.5 Software Infrastructure..... 7
2.5.1 Pilot Infrastructure 7
2.5.2 Demonstrators.....7
2.5.2.1 PILIN Web Handle Management Tool.....7
2.5.2.2 PILIN FRBR Tool.....7
2.5.2.3 Transfer Resource Demonstrator.....7
2.5.2.4 Appropriate Copy and Multiple Resolution.....7
2.5.2.5 Reverse Lookup Service.....7
2.5.3 Toolkits.....7
2.5.3.1 JAHDL.....7

3 Stakeholder and Stakeholder Engagement..... 8
3.1 Stakeholder Projects..... 8
3.2 Other Stakeholder Engagement Summary..... 8

4 Planned and Actual Milestones and Completion Dates..... 9

5 PILIN Interaction with other projects..... 13

6 Evaluation of Achievements..... 16

7 Lessons Learned..... 20

8 Sustainability and Future Activity Recommendations..... 23

9 Budget and Financial Arrangements..... 27

10 Project Participants..... 28

11 Resources..... 32

12 Final Comment..... 34

Appendix A: Stakeholder Projects..... 35
13.1 ARCHER with ARROW: Higher Education Research..... 35
13.2 State Library NSW..... 49
13.3 The Learning Federation (TLF) Schools Sector..... 55
13.4 University of Queensland: Integrating Identifiers into Annotation Harvesting..... 72
13.5 LORN (in conjunction with education au) VET Sector..... 84

Appendix B: Other Stakeholder Engagement Summary..... 99

1 Overview of Project

1.1 Background

The PILIN project (Persistent Identifier and Linking Infrastructure) was identified as an important national initiative in the third quarter of 2006. It was approved as a sub project of the ARROW II (Australian Research Repositories on line to the world) suite of activities in September 2006.

As various projects matured, including FRODO and MERRI, there was a growing realisation that sustainable identifier infrastructure was required to deal with the vast amount of digital assets being produced and stored within universities. This was and is a particular challenge for e-research communities where massive amounts of data are being generated without any means of managing this data over any length of time.

Many organisations, including the National Library of Australia, had been experimenting with persistent identifiers for some time. However there was no consensus or consistency across the educational and cultural/collecting agencies sectors. The PILIN project was an opportunity to comprehensively test the needs and parameters of identifier infrastructure as well as to establish the foundations of a future national service.

Prior to project commencement it was decided to use the Handle technology (http://www.handle.net) as core software while at the same time approaching the project from a technology neutral stance as far as was possible.

Despite its ARROW base the PILIN project was set up to have diverse stakeholders across all education and into cultural organisations/collecting agencies.

1.1.1 Why identifiers?

Vast amounts of digital assets are now being produced and stored by research and education institutions. There is an obvious need to manage, discover and access these resources over time. This is a particular challenge for e-Research communities where massive amounts of data are being generated by scientific instruments.

As repository projects have matured, it has become clear that identifiers are crucial for managing large numbers of digital assets over time. Identifiers help manage resource repositories by separating the identification of resources (their names) from the location of resources (their addresses).

Persistent identifiers can relate to web resources (e.g. content in an institutional repository), e-research and e-science resources (e.g. instruments, computational services, data sets), and semantic web abstractions (e.g. concepts in an ontology).

1.1.2 Why identifier infrastructure?

Identifier infrastructure enables management of the full lifecycle of a resource (from creation

to archiving) by guaranteeing:

- **global uniqueness** of identifiers through a registration process,
- a **persistent relationship** between an identifier and the same resource over time, and
- **reliable services** for managing and using identifiers over time.

Identifier infrastructure with these characteristics supports:

- **Access to resources over time.** Identifier infrastructure can support access to resources as they are moved from one repository or data store to another.
- **Discovery.** Persistent identifiers provide the link between the discovery of resources from federated repositories and access to those resources. The metadata returned from a discovery service can contain an identifier that is resolved to the location of the resource at the time of discovery.
- **Enhanced resource management.** Identifiers can be used to model complex relationships between resources and make it possible to manage compound resources, distributed copies of a resource, distinguish between different versions of a resource, and help decide which is the most appropriate copy or version for a user.
- **Rights management.** Identifiers can be used to associate rights information with resources. They can also be used to track the relationship between a resource and derived works.

1.1.3 Why shared identifier management infrastructure?

Shared infrastructure provides the redundancy (that individual users of Handle or other identifiers could not achieve). This is essential for guaranteeing **persistence** of identifiers and identifier services over time. It is also easier to **scale** shared infrastructure to the large number of identifiers and resolution requests required by the e-Research community.

Shared infrastructure also assists **movement** of identifiers and resources throughout the lifecycle of a resource. For example, data from a scientific instrument might be managed for a short time in a repository associated with the instrument. After that time it is transferred to a repository managed by the project that requested the data. When the data has been analysed it is archived in another repository for later analysis and replicated to an international repository.

Although technically the existing identifier infrastructure will support this movement of resources, the community needs **tools** to simplify the generation and registration of the original identifier and location, and simplify updating the location each time the resource moves. At any point in this lifecycle, the resources, metadata and identifiers can be separately managed.

Finally, the technology isn't enough. Persistence is not just a technology issue. It requires long-term **governance** and **policy** support at institution, sector and global levels. By developing a shared infrastructure, the Australian education and research sector can influence the future of identifier infrastructure at all these levels.

With this starting point the PILIN Project set about capturing the critical use cases and broad community requirements and adding value to the pre existing Handle software.

1.2 Aims and Objectives

The overarching objective of the PILIN Project was to:

> Strengthen Australia's ability to use global identifier infrastructure

Within that the more specific objectives were to:

- **Support adoption and use of persistent identifiers and shared persistent identifier management services by the project stakeholders.**
- **Plan for a sustainable, shared identifier management infrastructure that enables persistence of identifiers and associated services over archival lengths of time.**

These objectives were underscored by a desire to bring together a diverse cross sectoral community of stakeholders including:

- University Research Repositories (ARROW, APSR, RUBRIC, Digital Thesis, iSpheres)
- University e learning (MELCOE, USQ)
- University e research (DART/ARCHER, UQ, VERSI)
- Vocational Education and Training (LORN and EEG)
- School Education (TLF)
- Transport industry learning vertical (TALC/TILIS)
- Library communities (NLA and SLNSW)
- Others: ADL Labs, FRED, MAMS / RAMP, FRODO and MERRI, Dictionary of Sydney

1.3 Project Governance

The PILIN project was a pilot or demonstrator one. It was not developing sustainable services during 2007. The governance of the project fell under the normal ARROW Project arrangements with some additional pillars:

- The project leader, Kerry Blinco reported to the regular ARROW Management meetings and was the ultimate responsible person for the project
- The initial project patron Neil Mclean ensured the PILIN project fitted into overall educational infrastructure needs, Neil retired during the year.
- Regular reports were provided to DEST via ARROW (David Groenewegen)
- A steering committee chaired by Alan Smith from USQ, the project sponsor/partner/managing agent ensured the project kept on track and delivering as per the project plan
- A stakeholder group ensured that the diverse needs of the communities were heard and consolidated
- Within the available project funds, a small number of Extension Projects were developed

to allow further developments/testing of the toolkits and web services architecture

- The PILIN project team met weekly by teleconference and by face to face meetings approximately every 6 weeks
- The PILIN project team worked closely with aligned projects such as FRED and RUBRIC and personnel were encouraged to cross fertilise between the various projects
- A risk assessment was undertaken at the commencement of the project.

PILIN Risk Management

Risk	Current situation	Options to mitigate	Notes
Project unable to complete on time and on budget	Must be complete by 31 December 2007 and no avenue for additional funds. On budget and on plan at end of 2006	• Tight project management • Limit boundaries of scope of work	May be more an issue of what can be achieved by end of 2007
Project staff do not complete contracts	All staff recruited and operational	• All staff signed up until end of 2007 but no penalty clauses for early departures • Good spread of expertise across team so there is potential for coverage for emergencies	Could lead to non completion of work
Project distracted by Administrative issues	Managing an innovative and short term project through the University administration systems is a major risk for the project	• Establish work arounds to meet deadlines • Keep in constant contact with all impacted people on the project	This could be worthy of a minor paper at the end of the project recommending some strategies for future projects
Space for Strategic Leadership	Kerry Blinco and Nigel Ward provide strategic leadership for PILIN but their time is limited due to numerous other accountabilities and in particular their time for document review is restricted	• Recruitment of two Projects coordinators/communicators based at USQ to take on some duties/provide support for Kerry/Nigel • Devolution of some strategic tasks to other contractors eg Dennis Macnamara to manage CNRI liaison	Could lead to project not grasping all possibilities
Stakeholder expectations, (this is only pilot fabric)	Project is articulate about limited scope but expectations are high and interest strong and some	• Need for careful management of stakeholder involvement • Need for articulation of what could be done in 2008	**This is a critical risk**

	stakeholders have live up and running parallel projects		
Diverse nature of PILIN stakeholder communities and ARROW sensitivities	This project operates beyond the normal ambit of ARROW and across all learning and research communities and to some extent cultural orgnaisations	• Careful management of stakeholder engagement through the stakeholder group • Strong communication management by the Information Manager • Careful reporting to ARROW management group to ensure comfort with a project which sits a bit outside normal operations	Important political management issues here
Use case extraction	This project has potential to be over complicated with multiple angles and it may be difficult to analyse and articulate succinct use cases	• Project support for the business analyst to synthesise and prioritise use cases	Need to balance academic and business focuses
Project Scope and methodology	Project has an open ended scope to some extent in terms of what can be achieved in 2007 and for what spread of stakeholders	• Project methodology is an ***Agile*** one which allows the project to get done whatever can be done in the elapsed time and resources available • However the ***Agile*** methodology itself is a project risk in terms of defining outcomes clearly	Cross reference this risk with stakeholder expectations
Articulation to wider audience especially decision makers and funders of follow up work	As part of the project it is important to identify the source for turning the pilot into a real infrastructure and therefore to articulate the benefits and value of such an infrastructure in clear simple terms	• 2 of project team are noted translators of the complex and technical to the practical • Project accountability includes identification of future funding source	Vital to ensure that the 2007 investment in PILIN is leveraged

2 Project Outputs

All documents referenced in the Closure Report (as well as the Closure Report itself) can be found at www.pilin.net.au during 2008. However all documents have a handle reference for perpetuity.

Document	Location
2.1 Abstract Model for Persistent Identifiers - Ontology	• Ontology for Identifiers and Identifier Services http://resolver.net.au/hdl/102.100.272/G9JR4TLQH • PILIN Ontology Summary http://resolver.net.au/hdl/102.100.272/T9G74WJQH
2.2 Community Requirements	https://www.pilin.net.au/Project_Documents/Community_Requirements/Requirements.htm
2.3 Advice on Using Identifiers	
2.3.1 Internal Policy Documents	*https://www.pilin.net.au/Project_Documents/Policy_Documents/Policy_Docs.htm* • Handle Suffix Format for the PILIN Project http://resolver.net.au/hdl/102.100.272/Y35XYS0QH • Citation of Handles within PILIN Documentation http://resolver.net.au/hdl/102.100.272/R67T0T0QH • Identifier Association Policy for the PILIN Project http://resolver.net.au/hdl/102.100.272/3SYFCJ3QH • Authority Metadata Policy for the PILIN Project http://resolver.net.au/hdl/102.100.272/8RBZF0DQH • Version Control Policy for the PILIN Project http://resolver.net.au/hdl/102.100.272/KD633VNQH PILIN Filename Policy http://resolver.net.au/hdl/102.100.272/2C6K4VNQH
2.3.2 Guidelines and Considerations	https://www.pilin.net.au/Project_Documents/Community_Guidelines/Guidelines.htm • Meaningfulness of Labels in Identifiers http://resolver.net.au/hdl/102.100.272/D6N8F0DQH • Identifier Association Guidelines http://resolver.net.au/hdl/102.100.272/WBNMH9DQH • Considerations for Ownership of Identifier Management Systems http://resolver.net.au/hdl/102.100.272/461BL3DQH • Considerations for Managing Contexts http://resolver.net.au/hdl/102.100.272/N8R5K6DQH • Using URLs as Persistent Identifiers http://resolver.net.au/hdl/102.100.272/N8R5K6DQH • Form of Labels http://resolver.net.au/hdl/102.100.272/0HJ9X8JQH Identifier Service Guidelines http://resolver.net.au/hdl/102.100.272/1KKBLPDQH • Persistence of Identifier Guidelines http://resolver.net.au/hdl/102.100.272/V89DC0DQH
2.3.3 PILIN Glossary	https://www.pilin.net.au/Project_Documents/Glossary.htm

	http://resolver.net.au/hdl/102.100.272/HHYMV8JQH
2.3.4 Identifier Policy FAQ	https://www.pilin.net.au/Project_Documents/Policy_Documents/Identifier_Policy_FAQ.htm http://resolver.net.au/hdl/102.100.272/PFKJT1FQH
2.4 Service Oriented Models	https://www.pilin.net.au/PILIN_SUM.htm

All documents listed in 2.1, 2.2, 2.3 and 2.4 are available as one file at:
http://resolver.net.au/hdl/102.100.272/TDT3W6PQH

2.5 Software Infrastructure	
2.5.1 Pilot Infrastructure	(resolver.net.au)
2.5.2 Demonstrators	
2.5.2.1 PILIN Web Handle Management Tool	https://www.pilin.net.au/PILIN_Implementations/About_PWHMT.htm
2.5.2.2 PILIN FRBR Tool	https://www.pilin.net.au/PILIN_Implementations/About_FRBR.htm
2.5.2.3 Transfer Resource Demonstrator	https://www.pilin.net.au/PILIN_Implementations/
2.5.2.4 Appropriate Copy and Multiple Resolution	This is a FRED output in collaboration with PILIN and is located on the FRED website at: http://fred.usq.edu.au
2.5.2.5 Reverse Lookup Service	https://www.pilin.net.au/PILIN_Implementations/
2.5.3 Toolkits	
2.5.3.1 JAHDL	https://www.pilin.net.au/PILIN_Implementations/About_JADHL.htm

All software listed in section 2.5 is available at
https://www.pilin.net.au/PILIN_Implementations/Implementations.htm

3 Stakeholder and Stakeholder Engagement

3.1 Stakeholder Projects

Stakeholder	Project
The Learning Federation (TLF)	Appropriate Copy (with FRED project)
Learning Object Repository Network (LORN)	Relationship Service
State Library of New South Wales (SLNSW)	Digital Asset Management
ARCHER	Lifecycle of Data Objects
University of Queensland (UQ)/DART	Integrating Identifiers with Annotation Harvesting, Aggregation and Searching

A full report on each project can be found in Appendix A

3.2 Other Stakeholder Engagement Summary

The following bodies also participated in the PILIN project as stakeholders:

- APSR http://www.apsr.edu.au/
- ARROW http://arrow.edu.au
- Cooperative Research Centre for Spatial Information http://www.crcsi.com.au/
- Dictionary of Sydney http://www.dictionaryofsydney.org/www/html/7-home-page.asp
- Education.au http://www.educationau.edu.au/
- FRED
- Howard Florey Institute (HFI, Melbourne Uni) http://www.florey.edu.au/ and Arcitecta Pty Ltd, http://www.arcitecta.com as the HFI (Vendor)
- I Spheres http://www.ispheres.org/
- National Library of Australia (NLA) http://www.nla.gov.au/
- Neurosciences Research Group (Medicine, Nursing and Health Sciences, Monash University) http://www.med.monash.edu.au/medicine/mmc/neurosci.html
- RIDIR http://www.hull.ac.uk/ridir/
- RUBRIC http://www.rubric.edu.au/
- TILIS http://www.talc.com.au/TILIS/tabid/220/Default.aspx

4 Planned and Actual Milestones and Completion Dates

A number of distinct outputs were identified in the PILIN project plan developed in mid 2006. Following the initial appointment of some project team members, timelines were set in late 2006 for deliverables during 2007. As the project progressed, milestones and deliverables were altered to meet the emerging circumstances of the project. Most of the alterations were logistical rather than substantial changes of direction.

One significant change was that the Identifier Information Model or Ontology was developed. This deliverable was not specifically articulated in the original Project Plan, but was seen as critical for shaping the other outputs. At the commencement of the project in January 2007 when the team was fully appointed, it was felt that the Ontology would be produced prior to the completion of other policy guidelines and indeed software. As the project progressed it was realised that a concurrent not a sequential approach was needed. The ontology was developed in tandem with other outputs rather than as a precursor to them. In fact before the project commenced, this output was not articulated specifically as a deliverable.

Deliverable milestones in the project plan was also adapted in early 2007 so that outputs could be "released" to stakeholders, via the stakeholder wiki, prior to being released to the public PILIN website. In the project plan it had been felt that the products may be limited to just stakeholders "Access to the shared pilot infrastructure will only be made available to project stakeholders". During the project it was felt that while not promoting products, the outputs should as far as possible be available from the public website. As completed documents and software were made accessible in that way.

Project Outputs in Project Plan

- Best practice and policy guides for the use of persistent identifiers in Australian e-learning, e-research, and e-science communities.
- Use cases describing community requirements for identifiers and business process analysis relating to these use cases.
- E-Framework representations of persistent identifier management services that support the business requirements for identifiers.
- A pilot shared persistent identifier management infrastructure usable by the project stakeholders over the lifetime of the project. The pilot infrastructure will include services for creating, accessing and managing persistent digital identifiers over their lifetime. The pilot infrastructure will interoperate with other DEST funded systemic infrastructure. The development phase of the pilot will use an agile development methodology that will allow the inclusion of value-added services for managing resources using persistent identifiers to be included in the development program if resources permit.
- Software tools to help applications use the shared persistent identifier infrastructure more easily.
- Report on options and proposals for sustaining, supporting (including outreach) and governing shared persistent identifier management infrastructure.
- Articulation of the business value to education of shared persistent identifier infrastructure.

4.1 Milestones planned and actual

Deliverable	Planned Timeline	Actual Timeline	Notes
Ontology	Q1 2007	December 2007	Even at December 2007 this output is as much a work in progress as it is a finished product
Best practice and policy guides	Q1 2007	During the year as they emerged, from July to December	Most documents have reached a finished work status by December 2007. Some are notated as being "unfinished works" or outputs containing "open questions"
Documented community requirements	Q1 2007	Most in Q1 and Q2	These were released in Q3
e-Framework representations	Q2 2007	First release in Q2 but not finalised until Q4	While the e- Framework draft documents aided software development they were not really substantially completed until Q4 (and then as unfinished works) This was connected to the Ontology development
A "pilot" shared persistent identifier management infrastructure	• Release 1 Q2 2007 • Release 2 Q3 2007 • Release 3 Q4 2007	Q1 Q2	The infrastructure was released ahead of plan and a second release was done in Q2. In Q4 full documentation was released
Open Source software toolkits	• Release 1 Q3 2007 • Release 2 Q4 2007	Q2 Q3 And Q4	Toolkits were being demanded by stakeholders very early in the project so they were made available earlier than the project plan under a Mozilla open source software licence There were, in the end, a number of distinct toolkits (see Product listing) each of which had one or more releases. Full documentation for the toolkits was not released

			until Q4
Sustainability report	Q4 2007	Release 1 Q2 2007 Release 2 Q4 2007	A report was required for national budget planning in April and this was completed. This is not a public document Later in the year a paper was prepared jointly with the NLA for funding a future sustainable service. This is not a public document
Articulation of the business value (see Sustainability report)	Q4 2007	Release 1 Q2 2007 Release 2 Q4 2007	While the sustainability report and the articulation of the business value were in fact bundled as one, there was consistent and concerted articulation of the business value of persistent identifiers. All major stakeholders concurred that persistent identifiers were a crucial component of any repository and search and access service.

4.2 Project Methodology

The project team were geographically dispersed, with team members in Melbourne, Sydney, Brisbane and Toowoomba as well as project advisers/mentors in Canberra and the UK and USA.

The project used all available forms of communications, including a project team wiki for thought and document development. The team met weekly by teleconference and 6 weekly face to face.

Outputs from the project were first released via the stakeholder wiki and feedback from there was synthesised before subsequent releases and releases to the public website. Two major stakeholder meetings were held in March and November. This allowed concerted extraction of stakeholder requirements but also extensive knowledge sharing across the diverse stakeholders. At the same time all stakeholders were contacted and followed up on a regular basis after the initial use cases and requirements were extracted. In a number of cases the PILIN project worked in tandem with a live business service (such as the SL NSW Digital Asset Management activity and the LORN Project) and the PILIN software was to be used in these live operations. To date the only live business use of a PILIN related output (the FRED

developed Appropriate Copy service) has been with The Learning Federation

The project methodology also included the funding of a number of extension projects described in section 3.2

Apart from release to the public website few other outreach activities beyond the stakeholders was attempted. For instance the PILIN software was not released via SourceForge. However, PILIN team members did present on identifier issues and PILIN outputs at a number of conferences and events across Australia including; Clever Collections, IDEA 2007, eResearch Australasia 2007, MERRI at USQ Roadshow.

The overall approach to the development of the PILIN products was an agile one. This went beyond just an agile methodology for the software development. The policy development was similarly agile and in fact the software and policy development were intertwined and mutually reinforcing. The project Business Analyst was embedded with the software development team and was also jointly responsible for policy development. The interrelated technical and the policy issues were thus addressed cohesively. In this way the project could quickly respond to emerging issues and reorder priorities and milestones to achieve the desired outcomes albeit via a different pathway. This approach allowed what proved to be a more complex task than initially conceived to be mostly completed within a very tight timeframe.

5 PILIN Interaction with other projects

During the course of the project, the PILIN team made contact with a number of non-stakeholder projects working directory or indirectly with identifiers. Some of these are also cross referenced in sections 3.1 or 3.2.

RIDIR (United Kingdom) www.hull.ac.uk/ridir

The RIDIR project (Resourcing Identifier Interoperability for Repositories) is a project funded under the auspices of the Joint Information Systems Committee 'Repositories and Preservation' Programme. It is investigating the requirements for, and benefits of, the clear use of persistent identifiers in order to facilitate interoperability between digital repositories of different types.

Dennis Macnamara met with Hugh Look and Chris Aware from the RIDIR Project in April. The RIDIR team were given access to the PILIN Stakeholder wiki which included Use Cases gathered by the PILIN Business Analyst, Nick Nicholas. The RIDIR project included a number of the PILIN Use Cases in their RIDIR Focus Group Report[1]

"The second set of use cases is abstracted from the work of the PILIN Project in Australia. This is a relatively small subset of their use case work but represents the ones that seem possibly applicable to RIDIR's brief."

VIF Versions Identification Framework http://www.lse.ac.uk/library/versions/

Versions of Eprints - user Requirements Study and Investigation Of the Need for Standards

The VERSIONS Project addresses the issues and uncertainties relating to versions of academic papers in digital repositories, with a focus on the discipline of economics.

Frances Shipsey, from LSE (London School of Economics) and Project Manager of VIF also attended the meeting in London with Dennis Macnamara and the RIDIR project. This project has a life cycle running a fair bit behind the PILIN project so not much further interaction has occurred, but this will be an important project for the PILIN extension activities and the proposed national service to keep on the radar.

CNRI http://www.cnri.reston.va.us/

CNRI (Corporation for National Research Initiatives), based in Reston, Virginia, USA focuses on strategic development of network-based information technologies, providing leadership and funding for research and development of the National Information

1 , University of Hull and rightscom , RIDIR Focus Group Report September 2007

http://www.hull.ac.uk/ridir/Documents/workshopreportv10.pdf

Infrastructure. CNRI developed the Handle System used by the PILIN project. They also provided valuable feedback on the services and policy documents developed by PILIN.

PILIN team members visited CNRI and remained in regular contact with key CNRI Handle staff including Larry Lannom. It is proposed that Larry come to Australia in April 2008 to build the relationship between CNRI and the new national persistent identifier service in Australia

FRED

THE FRED Project (Federated Repositories for Education) aims to support deployment of repository federations in Australian education and training communities. It is documenting generic service-oriented models and producing software toolkits that support development of repository federations. The FRED project is sponsored by the Australian Commonwealth Department of Education, Science and Training under the Framework for Open Learning Programme.

The FRED project had an ongoing and productive interaction with the PILIN project. This was aided some key staff being involved in both projects. The FRED / PILIN interaction examined the question “How does OpenURL relate to persistent identifiers?”. The answer resulted in the FRED project specifying, documenting and producing an Appropriate Copy service that used OpenURL as a front-end, and the Handle system as a back-end. The software is available under an Open Source license and has been demonstrated to PILIN stakeholders.

DART/ARCHER http://dart.edu.au/ (this project became a stakeholder written up in section 3.2)

The DART/ARCHER project is a proof-of-concept project to develop tools to support the new collaborative research infrastructure. The project aims to enable researchers and reviewers to access original and analysed data, collaborate around the creation of research outputs and stored publications, plus add content, annotations and notes. It will also look at the collection of large datasets, including the remote control and automated data collection.

PILIN and DART ARCHER worked closely together in 2007, aided and abetted by both having a Monash university presence. PILIN funded an Extension project with ARCHER and the PILIN senior software developer moved on to work with ARCHER just before the completion of the PILIN project.

APSR http://www.apsr.edu.au/index.html (also written up in stakeholder section 3.1)

The APSR (Australian Partnership for Sustainable Repositories) Project aims to establish a centre of excellence for the management of scholarly assets in digital format.

APSR staff kept in close contact with PILIN, attending both stakeholder meetings and providing valuable feedback on repository needs.

MAMS http://www.melcoe.mq.edu.au/projects/MAMS/

Macquarie University is the lead University on the Meta Access Management System (MAMS) Project. The project allows for the integration of multiple solutions to managing authentication, authorisation and identities, together with common services for digital rights, search services and metadata management.

MAMS staff attended the first PILIN stakeholder meeting and provided valuable input into PILIN use cases early in the project.

RAMP http://www.melcoe.mq.edu.au/projects/RAMP/index.htm

The Research Activityflow and Middleware Priorities (RAMP) project seeks to improve national research effectiveness by addressing two of the most challenging components of the DEST/JISC E-Framework for Education and Research and the DEST Accessibility Framework – the areas of people-oriented workflows for research processes, and open standards authorisation for protected repositories.

Ramp staff attended the first PILIN stakeholder meeting and provided clarification help in terms of the interface between identification of objects and identification of people particularly in a workflow sense.

6 Evaluation of Achievements

The PILIN project achieved significant results especially for a start from scratch operation in a short time frame. The diverse team came together really quickly and the project was fortunate to recruit a number of experienced and highly skilled personnel. The team stayed largely in tact for the whole project which again was an excellent result for such a fixed term project.

The achievements of the PILIN project can be categorised as follows:

1. Software toolkits and outputs
2. Policy: guidelines and considerations
3. Sector/Stakeholder impact
4. Persistent Identifier impact

6.1 Software toolkits and outputs

These have been listed and linked to in Section 2.5. They represent a significant addition to persistent identifier software resources. They make it much easier, simpler and user friendly to use the Handle software. They also add more functionality in terms of batch handling, FRBR relationship services, and reverse look up. With this additional software available for users it is anticipated that more Australian entities will take up the use of Handles. The software is all available through an open source software licence.

The software however has been developed in a project and generic environment. PILIN software has not yet been used in a live business operation. The Learning Federation Appropriate Copy service for the schools sector is a live business operation but that was using a FRED software output albeit within a PILIN extension project context. While there is every reason to believe that the PILIN software is of the highest quality, it still needs further testing of its robustness. At the same time there will always be a need for porting to specific operating environments. The Learning Federation confirmed this in their implementation as have the National Library of Australia (NLA) when PILIN software was installed on their servers.

At the same time it is recognised that PILIN only developed some of the software toolkits that will be needed by Australian repository services. It was not possible to develop all software identified in the use case gathering (see section 2.2). Further work that can be done can be found in section 8 of this report.

6.2 Policy: Guidelines and considerations

The list of policy products found in section 2.3 is extensive. The project developed policy for itself in terms of identifier management and also developed a set of community guidelines and considerations. The internal policy documents have been made publicly available to aid others in their policy development but are not intended to be prescriptive policy for other organisations. The community guidelines and considerations have been written for other

organisations. They represent the combined wisdom of the PILIN project in 2007 and are not therefore intended to do more than inform any specific organisational policy development.

The work has added depth and coverage to the field and provided support and guidance for the repository communities in Australia and beyond. The individual guidelines are backed by a glossary and an information model or ontology as well as a policy FAQ. Feedback on the guidelines has been very positive and they have already been used to frame structures and frameworks for repository implementations.

However, partly due to the project logistics, the documentation has not been fully road tested across the communities. Some feedback has been received and incorporated into revised versions but there is more work to be done to fully develop the guidelines into fully and broadly road tested documents. Some of the guidelines contain open questions and areas of debate that have not been resolved by the project. Where possible, these open questions are noted in the guidelines documents to help stimulate resolution in the future. Other guidelines documents are still works in progress at completion of 2007. These are still released by the project, but are marked as "unfinished works". Moreover there are persistent identifier policy and governance areas not covered by the PILIN documents. These are noted in section 8 of this report.

Nevertheless, the body of work represents a substantial effort and will continue to inform educational, research and collecting communities for a number of years.

6.3 Sector/Stakeholder impact

The PILIN project gathered together a broad cross section of stakeholders (see section 3). It has impacted on those stakeholders in a number of critical ways:

- Gathering and documenting use cases (see section 2.2)
- Providing significant input into stakeholder thinking about the use of identifiers
- Participating in the debate on the use of identifiers in Australia
- Providing software for use by stakeholders
- Providing policy guidelines for stakeholders
- Giving stakeholders an opportunity to participate in the discussions and debates around persistent identifier use
- Providing stakeholders with easier access to Handle persistent identifier software
- Developing software for use by stakeholders in live business projects (eg TLF)
- Bringing together persistent identifier expertise from across the sectors and forging an Australian community of practice for persistent identifiers
- Synthesising the similarity of needs from divergent sectors (eg e-research and Vocational e learning)
- The funding of various stakeholder extension projects from the PILIN budget allowed a number of stakeholders to advance their projects and business activities while at the same time allowing the PILIN project to gain valuable insights into the application of identifier

software and policies to real business ventures. Each of the extension projects have been described and written up (see section 3.2)

The PILIN project has therefore had a significant beneficial impact on its stakeholders. Some had already started work on identifiers but the PILIN project gave that work a huge boost. Others had been waiting for such a project as PILIN before putting their toes in the water. Overall the PILIN project provided a huge boost to persistent identifier take up amongst it stakeholders.

However there is always more that could have been done and in some cases it was not possible to align PILIN rhythms and logistics with stakeholder needs and timetables (eg TALC/TILIS). In other cases there were not enough resources to pursue some stakeholders in order to do real business together (eg I Spheres). Some stakeholders were not quite ready to fully utilise PILIN outputs during the lifecycle of the PILIN project (eg SLNSW).

Having laid the foundations PILIN is expectant that the stakeholder impact will continue to bear fruit in 2008 and beyond. Many of the stakeholders are planning to implement persistent identifiers in 2008 using PILIN software toolkits/Handle software.

6.4 Persistent Identifier Impact

The PILIN project represents one of the most substantial efforts within Australia on the use of persistent identifiers. The project has undertaken a number of initiatives to ensure maximum impact including:

- Fully documenting and publicly releasing all outputs and thinking (even when not finalised)
- Presenting papers at a number of conferences and events across Australia
- Co authoring submissions and strongly lobbying for the funding of a national persistent identifier service in Australia (including the identification of and advocacy for a service provider)
- Confirming the absolute need for persistent identifiers (and Handle as an appropriate current technology)
- Enabling the establishment of a National Persistent Identifier Service
- Liaising with CNRI, the developers of Handle, in the USA to ensure that the Australian service provider can obtain a suitable licensing arrangement and have a seat on the global board of governance for Handle.
- Liaising and sharing with other global projects, RIDIR (see section 5) in the UK for instance to ensure the work in Australia gains global recognition as well as supporting other projects. The RIDIR project in the UK has used the PILIN Use cases in its work.

However, the PILIN project is only a starting point. By the end of the Project the PILIN team still had some unanswered questions; in fact in some ways the issues around identifier governance and the meaning of persistence grew more complex as the year progressed. There was also a remaining need to provide simpler synthesise documents for wider repository user consumption. While those in the know had no doubt about the absolute necessity for

persistent identifiers there was still a need to “educate” the wider community.

There is still much outreach work to be done. There was also no time to write journal articles about the project and the development of identifier shared infrastructure. This is important to undertake in order to give further credence to the persistent identifier developments.

But overall the PILIN project has had a major impact on the use of persistent identifiers in Australia and to some extent globally. One senior practitioner suggested “Australian organisations have been playing around with persistent identifiers for quite a few years but we have never really made sense of them. PILIN has really brought things together and provided as path forward”. The overall PILIN objective of “**Strengthening Australia's ability to use global identifier infrastructure”** has indeed been met.

7 Lessons Learned

There were a number of key things learned from undertaking the project. Some were to be expected, some were less obvious. Many of these lessons were discussed with the stakeholders including at the meeting on 21 November.

7.1 Project governance

- It was critical that the project had robust project governance. This was achieved through a small senior level group chaired by the project sponsor and clear reporting lines to the funding source. The Steering Committee also insisted on a rigorous risk analysis at the commencement of the project.

7.2 Project management and logistics

- The PILIN project was a geographically dispersed one with staff in Melbourne, Sydney, Brisbane and Toowoomba. We learned that such geographic diversity could be managed with tight project management and regular contact through multiple means. The geographic diversity, including the options to tap into global expertise, greatly enhanced the firepower of the project. However it also added to the project complexity.
- The PILIN team contained a diversity of specialists and it may have been better to spend more time up front ensuring everybody shared the same vision and talked the same language.
- There was perhaps a lack of clarity early on in the project in terms of policy authoring and documentation. While this slowed down some outputs it was part and part of normal project development and in the end documentation was the better for having a number of inputs. The bundling of the business analysis and initial policy authoring activities, while resource straining, proved to be a critical success factor.
- More time spent in face to face meetings early in the project may have forestalled miscommunication issues or rather may have better shaped the project outputs planning.
- Commencing a project late in one year and ending on 31 December the following year presented serious challenges in terms of the traditional dead Christmas period from mid December to mid January and the end of financial year issues within universities. As far as is possible such start and finish times should be avoided.
- The project underestimated the administrative impact and burden on the host university, USQ. The volume of additional contracts and additional payment and invoice transactions placed a heavy load on the staff at USQ.
- The importance of strong but flexible project management was noted. The agile methodology used for both software development and policy development and indeed overall goal setting was a key strength of this project. However it required concomitant strong project management to ensure the project reached most of the intended outcomes. However sometimes, especially in the first half of the project, we were slower to release

draft products to stakeholders than we would have liked. There was a need to fine tune the internal and stakeholder review processes which was done for the second half of the project.

- The use of e framework approaches provided a good methodology for project software outputs. Nevertheless the service expressions around identifiers remain unfinished works at project completion.

7.3 Stakeholder engagement

- The importance of gaining support from a diverse range of stakeholders was noted and achieved.
- Getting out to stakeholders early in the project was a critical success factor for the project especially in terms of extracting use cases and building platforms for collaboration.
- Trying to adapt PILIN outputs to fit in with live business projects was challenging but essential. There was a need to acknowledge that the project stakeholders were not actually clients so we could not necessarily release software and policy guidelines exactly as needed or when needed by their timelines.
- There was probably a lack of recognition of the duality of interaction with stakeholders ie educative and use case extraction. While this is true of most projects it was probably not identified as such at project commencement.
- The importance of closing the loops for example by running a second stakeholder day towards the end of the project should be noted for other projects.
- The need to accept changes in stakeholder priorities and availability for cooperative work was perhaps underestimated by the project team. It should also be noted that stakeholders were at varying points of preparedness, in terms of resource availability, planning and staffing. Those who were less prepared in terms of planning were the least able to make most use of the PILIN engagement.
- The advantages of cross sectoral stakeholders was emphasised in this project. It added tremendous value to the project but it also added value to the individual sectors activities.
- The setting up of extension projects should probably have been done earlier in the project even though this would have been problematic in terms of having enough to go on at that time.
- Need for more inducement and incitement for stakeholders to participate in wiki type activities.

7.4 Project content

- It was a much more challenging project than the project team envisaged at project commencement. As the project progressed we became less sure about some of the key issues.

- The project team struggled to both deal with the complexities and keep it simple/user friendly at the same time.
- The development of an ontology for persistent identifiers became a whole project endeavour and was only completed at project end. In fact it was not really articulated at all in the project plan despite being a necessary activity. Overall the ontology or information modelling caused the most angst within the project especially coordinating the input from advisors in UK and USA with the work of the development team at Monash.
- The governance and policy issues for identifiers were more problematic than software development (but we suspected this).
- Stakeholder needs across the sectors, while having some differences, had a common core of needs making a national cross sectoral service a serious business proposition.
- Software development for a generic set of users needs some additional material to allow it to be deployed in specific situations, eg porting for Solaris operating system used at NL
- The ontology and policy guidelines may need a number of versions especially to make them accessible to various levels of repository users - users guide to information modelling.
- There are still some key policy areas to be addressed. For instance how long is persistence is clearly contextually based but it may be possible to develop community guidelines to aid specific context implementations. Persistence might be considered in terms of accessible/useable persistence and archival persistence.
- There were a number of additional services or software products identified during the project where there was not time to develop. These include:
    - Improved reporting and managing of Identifiers (track my identifiers)
    - Identifier Validation (identifier link-rot checker)
    - Digital Object Authentication (check-sum validation of identifier targets)
    - Hosted Handles server

8 Sustainability and Future Activity Recommendations

The PILIN project has made great advances in 2007 in the development of capability in Australia to successfully utilise persistent identifiers. However this work will come to nought without an ongoing national sustainable service to support future developments at the individual organisation level.

In the last decade a number of organisations in Australia have invested time and resources into developing persistent identifiers infrastructure for their repositories. However the field is littered with broken links and persistent identifiers not being persistent regardless of the identifier software used. This has a lot to do with policy and governance matters and less to do with software and technology, although a more user friendly access layer may make identifier use more ubiquitous and more lasting.

PILIN has established an Australian foundation for making use of available global identifier infrastructure. The PILIN project mainly worked with and extended Handle software infrastructure as it believed it to be the most flexible and malleable available. However the governance and policy underpinnings are software neutral and applicable to most identifier systems.

As part of its project objectives PILIN has made recommendations for a national sustainable service and identified and worked with the most likely provider of such a service, the National Library of Australia, NLA. At project closure it is probable that such a service can be funded from ANDS (the Australian National Data Service as part of the National Collaborative Research Infrastructure Strategy, NCRIS e research thrust). This service is likely to commence in the second half of 2008, or at least receive investment funding to commence development of such a service. As the monies for this are coming from e-research there is still a need to ensure that the National Persistent Identifier Service will be available across education and collecting agencies sectors.

In order to maintain momentum and prepare the ground for a national service, some funding has been secured for PILIN to extend its work into 2008, at least until the end of February and possibly until the end of June. This will be considered an extension phase and the initial PILIN project will formally close on 31 December 2007. The recommendations in this section of the closure report may in part then be addressed by the PILIN extension activities although much will be part of the work of an ongoing national service.

The PILIN project started making submissions on a sustainable national service as early as March 2007 in order to ensure there was the best possible chance of it coming to fruition. The following extracts from the initial submission for future identifier services illustrate the early thinking of the rationale for such a service:

> Persistent identifiers are a common need across educational and cultural sectors. Many institutions and organisations are looking to digitise their assets. Learning cultural and research resources are subject to more movement and more reuse. Establishing an Australian utility and linked services will offer economies of scale efficiencies to all sectors. The costs of each sector and sub sector independently establishing such utilities and services would be wasteful and prohibitive leading to non take up of persistent identifiers. An Australian focus on persistent identifiers will also encourage take up of a persistent identifier approach to doing business. This in turn will enable greater leverage and use of

> all cultural, learning and research assets and in many cases will drive further business opportunities. The success of the Australian creative economy will be dependent on many infrastructure investments. Persistent identifiers are a key jigsaw component of such an infrastructure.
>
> Persistent identifiers are a key component of the jigsaw to enable maximum leverage from research, learning resources, cultural artefacts and assorted data and information. The strength of the future Australian economy will be dependent on such leverage. Providing a national sustainable infrastructure service will be a huge impetus for sustainable growth and well being. A myriad of independent non-linked identifier developments across educational sectors and cultural/creative industries will result in lost opportunities and gross inefficiencies. Unlocking educational and cultural resources for maximum community and economic benefit through the persistent identifiers development will be a strategic and critical investment.

As the project progressed there was overwhelming endorsements from key stakeholders that a persistent identifier national service was an absolute necessity. The PILIN project sees the national service as an economic necessity for Australia's continuing success in a digital environment.

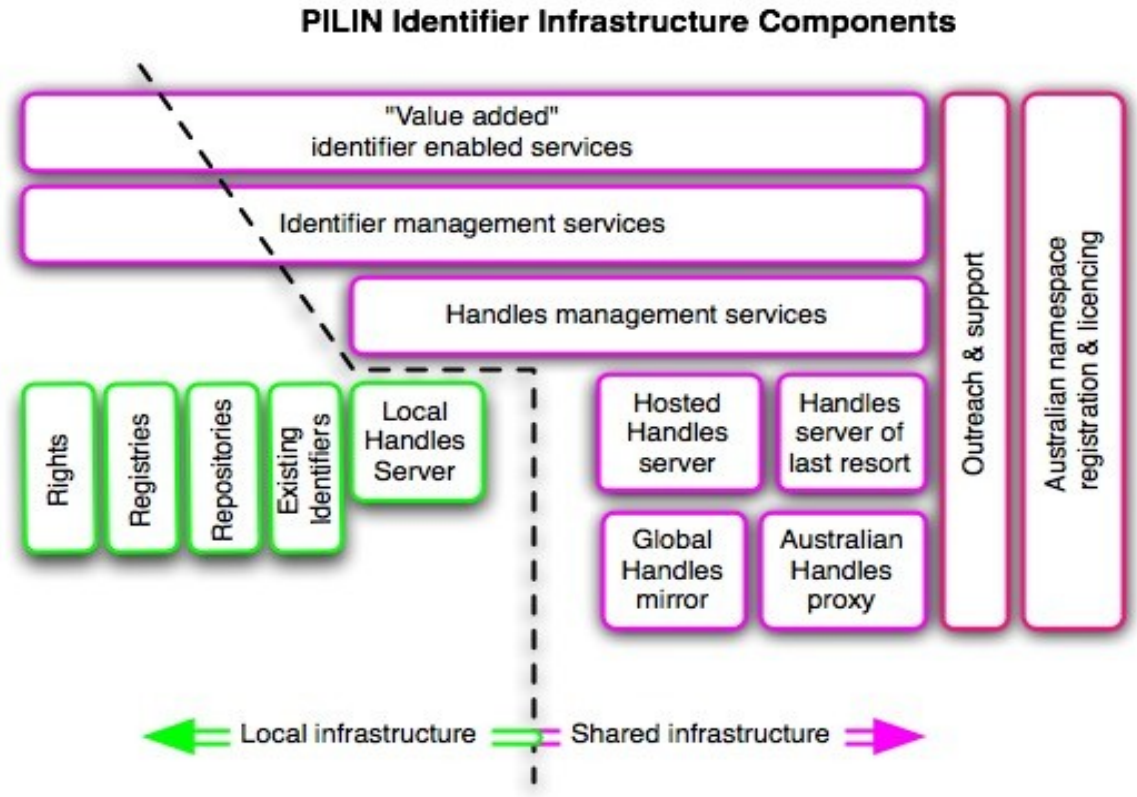
8.1 Recommendations for future activities

The PILIN project has identified a number of activities that could be undertaken to further develop Australia's ability to use global identifier infrastructure. Some of these activities may be undertaken as part of the transition to a national persistent identifier service. Some will be part of the work of that national service and some will be the province of future research and development projects.

Recommendations for future activities are listed below under specific categories of work.

8.1.1 Governance and frameworks

- A national service must be established to provide basic utility services for persistent identifiers
- The national service must be available across all education and collecting agency sectors regardless of the funding source for the service
- Significant outreach services are required to ensure repository capability to best use identifiers for persistence
- There needs to be a framework for central utility services and local repository and other services. The following diagram is seen as a starting point



- **In order to make optimum usage of persistent identifiers the following schema is suggested** for what local repositories need to do from a policy and process perspective:

  Model information
  Include identifiers in the modeling
  De-couple identifier management from information management

Then it will be possible to build information management services that leverage identifiers

8.1.2 Further policy and user requirements work

- Support for non-Handle identifier schemes next year (and migration services ie support for repositories that may wish to move from another identifier system to Handle)
- Bridging generic policy to specific implementation requirements of specific services eg a national e-Research service
- Users guide to identifier aware information modelling
- Policy guidelines on persistence, how long for useable, how long for archival
- Improved reporting and managing of Identifiers (track my identifiers)
- Identifier Validation (identifier link-rot checker)
- Digital Object Authentication (check-sum validation of identifier targets)
- Identifying further services that are needed to ensure optimum identifier implementations

8.1.3 Software development

- Making the existing software fully production strength especially for those services vital to the national utility service
- Developing software based on new use cases for migration services, reporting and managing identifiers, identifier validation and digital object authentication

8.1.4 Other

- Designing and developing a Hosted Handles server
- Developing a position paper on CNRI Handle licence arrangements and participation in Handle global governance arrangements
- Developing a transition plan for the new national utility service. The plan will need to describe dependencies on other services that are required for a viable persistent identifier service. The plan will also need to identify mechanisms for broad involvement of stakeholders in ongoing activities.

8.2 Final Comments

As the national service is rolled out new issues and needs will arise. The recommendations in this report are what are known at the end of 2007. Nevertheless they represent a solid base for future service development. It is pleasing to complete the PILIN project with high expectations that the work will be used in the development and production of a into a national utility service.

9 Budget and Financial Arrangements

A budget of $1.4 million dollars was allocated to the PILIN project as part of the ARROW II (Australian Research Repositories on line to the world) suite of activities in turn under the Australian Commonwealth Department of Education, Science and Training, DEST) Systemic Infrastructure Initiative. Project funds were to be expended between October 2006 and December 2007. Most team members were recruited in December 2006/January 2007.

The Project sponsor was University of Southern Queensland (USQ) and most of the PILIN team members were contracted directly by USQ. However the monies for the project were managed and accounted for out of ARROW central at Monash University. USQ invoiced Monash for its PILIN project expenses. Monash in turn reports to DEST.

The detailed budget plan and indicative outcome are as follows.

Item	**Budget $,000**	**Committed $,000**	**Expended $,000**	**Notes**
Total Development	**655**	**659**	**580**	
Software Developers	400	419	415	
UK /USA trainer/mentors	35	50	50	
Server /hosting/supplies	55	60	60	
Extension Projects	120	120	45	Costs of most extension activities within consultants and software developers costs
Contingency	45	10	10	Monash management fee
Policy	**745**	**795**	**788**	
Consultants/information manager	720	790	773	Consultants Include, business analysts, policy developers and project management
Events /Stakeholders	25	15	15	
Totals	**1400**	**1464**	**1368**	Final figures will be available after January reconciliation

10 Project Participants

PILIN Team

Kerry Blinco

Project Leader

Dennis Macnamara

Project Manager

Nigel Ward

Technical Director

Nick Nicholas

Business Analyst

Tina Reedman

Information Manager

Amit Chaudhary

Senior Systems Developer

Manish Saroha

Systems Developer

Xiaoming YU (Simon)

Systems Developer

Lyle Winton

Technical Advisor

USQ Support Staff

Ashley Lourey

Web Resources and Project Support Officer (Divisional Projects)

Cynthia Douglass

Project Officer

Donna Tilley

e-Framework and Standards Project Officer

Mentors and Key External Support

Alan Smith

USQ Project Sponsor

Katie Blake

ARROW Information mentor

Neil Dickson

ARROW Research Content Manager

David Groenewegen

ARROW Manager

Judith Pearce

NLA Library information mentor

Imran Khan

University Queensland Annotation

Simon Porter

DART Annotation Harvesting mentor

Dan Rehak

Technical Mentor

Peter Sefton

Software and systems mentor

Andrew Treloar

ARROW, DART and ARCHER conduit

Rob Wilson

Technical Mentor

Steering Committee

Alan Smith

USQ and Chair

David Groenewegen

ARROW

Judith Pearce

NLA

Andrew Treloar

DART

Kerry Blinco

PILIN Project Leader

Dennis Macnamara

PILIN Project Manager

Stakeholder contacts

NAME	Organisation
Steven Hayes	ACL, University of Sydney
Adrian Burton	APSR
Marcus Buckhorne	APSR
Cyrus Keong	ARCHER
Jason Lohrey	arcitecta pty. ltd
Neil Dickson	ARROW
Anthony Beitz	DART/ARCHER
Graham Reynolds	DEST
Tom Ruthven	Digital Thesis
Jerry Leeson	Education au
Geoff Hendrick	education.au
Owen O Neil	FLAG E standards/LORN
Tristan Gutsche	iSpheres
Neil Clarke	Monash University
David Groenewegen,	Monash University
Andrew Treloar	Monash University
Tony Boston	NLA
Debbie Campbell	NLA
Matthew Walker	NLA
Chi Nguyen	RAMP/MELCOE

Sue Craig	RUBRIC
Simon Handfield	SLNSW
Suzanne Moir	SLNSW
Marcel Chaloupka	TILIS
Preety Agarwal	TLF
Neil Killeen	University of Melbourne
Jane Hunter	University of Queensland
Anna Gerber	University of Queensland
Peter Sefton	USQ
Paul Davis	VERSI
Larry Lannom	CNRI
Hugh Look	Rigtscom
Chris Awre	University of Hull

11 Resources

At one level The PILIN project had the resources outlined in sections nine and ten of this report. However, more than the dollars, the project team, the stakeholders and the Steering Committee, the PILIN project was able to harness an array of physical and intellectual property that underpinned the success of the project.

The PILIN development team was housed at Monash University Library in the ARROW central unit. It used the physical and network facilities of Monash University and was able to benefit from the intellectual stimuli from aligned projects and the wider university environment. The project relied on Monash servers for software development and wiki communications.

The Information Manager was housed at the State Library NSW and benefited from both their infrastructure and the intellectual support. The State Library became a stakeholder in the PILIN Project and enthusiastically embraced the project outputs.

Melbourne University housed two of the PILIN team advisers and proved a valuable source of reference especially in the e- research aspects of identifier use.

The PILIN project was sponsored by University of Southern Queensland (USQ), and while no contractors were based at USQ, all were supported by a team of professional technical and administrative staff as well as advice and support from the project sponsor Alan Smith. USQ also provided technical infrastructure for the PILIN public website and software, ICE and TRAC, for both the website and for the wikis.

The emerging Link Affiliates group was also a useful support network for the PILIN project. Link Affiliates is the branding of the Department of Education, Science and Training (DEST) developed technical standards and interoperability capabilities. Link Affiliates are a distributed team, based at University of Southern Queensland and with partnerships at Monash University and the University of Melbourne. Link Affiliates participates in international standards development; manages key international partnerships and relationships and conducts standards based interoperability demonstration projects. Link Affiliates disseminate information about standards activities and participate in international standards development.

The PILIN project used the CNRI Handle software as its base. It developed improved functionality and additional services through Java based applications. Software used in the project included:

- Eclipse IDE v 3.2.1
- MySQL DB v 5.0
- SVN Repository v 1.4.2
- Glassfish Server v 2
- Java v 1.6
- Tortoise SVN GUI v 1.4.1
- TRAC v 0.9b1
- Handle Server v 6.2
- Luntbuild continuous Integration Tool v 1.3.6

- Toad for MySQL v 2.0.3
- Star UML v 5.0

Much of the project has been informed by previous work on standards and e framework http://www.e-framework.org/ .Two of the PILIN team brought extensive experience from these arenas and continued to operate in key standards forums during the lifecycle of the PILIN project. Other members of the team were also exposed to the issues and knowledge emanating from standards forums. Some members attended key global standards events during 2007.

Policy work was informed by the work in a number of other projects including:

- ARROW
- DART/ARCHER
- APSR
- FRED
- Merri
- FRODO
- ADL

The PILIN project then was undertaken on a solid but evolving foundation of global resources and good practice.

12 Final Comment

The PILIN project has completed. It has set the scene for a sustainable future service. At the time of writing this report a short PILIN extension activity has been approved as part of the transition to a future national service during 2008. PILIN extension activities will be reported on separately.

13 Appendix A: Stakeholder Projects

13.1 ARCHER with ARROW: Higher Education Research
http://ARCHER.edu.au/

pilin persistent identifier linking infrastructure

web: http://resolver.net.au/hdl/102.100.272/0N8J991QH

email: policy@pilin.net.au

To cite the *latest* version of this work use http://resolver.net.au/hdl/102.100.272/8FF47GNQH
To cite *this* version of this work, use http://resolver.net.au/hdl/102.100.272/8FF47GNQH

Stakeholder	Name of Project	Resources used in terms of time estimate of $	$ allocated out to stakeholder
ARCHER with ARROW	Procedures and workflows for the Provisioning of Persistent Identifiers for Datasets	Project scoped on 31 August by business analyst and DART/ARCHER personnel. $15Kfor PILIN team resource time and external costs	Nil

13.1.1 Description of Project and Objectives

ARCHER wishes to establish a role for persistent identifiers in the life cycle of e-research data objects, particularly as they cross the curation boundary between a collaboration space and a publication space. This will result in best practice recommendations, and integration with a persistent identifier infrastructure in the data management toolkits that ARCHER will release.

To realise this requires three components: first, a report from PILIN on the policy guidelines necessary to provide a persistence identifier infrastructure that involves multiple policy domains/institutional contexts (collaborators' home institutions, collaboration space vs. publication space vs. experiment validation space).

Second, integration of infrastructure for creating and managing persistent identifiers with the asset management tools used in the collaboration space. Identifiers for objects in the collaboration space are likely to be cited but not resolved, since their persistent association with an access point is conditional on the object being published (leaving the collaboration space). Therefore identifiers are anticipated to be used at this stage as placeholders. The manner and extent of integration was left open, and could potentially be as simple as metadata tagging a dataset with a Handle, instead of actually registering a Handle record (because the persistent identifiers are placeholders at this stage). PILIN's involvement is to

provide support for an abstract model of asset management. The specifics of asset management, and the particular integration of PILIN tools into the asset management, will be the responsibility of ARCHER subsequent to PILIN's work.

Third, support for populating persistent identifiers with the resolution data appropriate to the publication space (from ARCHER SRB to ARROW Fedora), when the data objects cross the publication boundary. ARCHER has in its project scope the importing and exporting of scientific data, but not the publishing of scientific data: publishing may be provided by ARROW. In other words, the final stages of the process may involve ARROW as a stakeholder.

The final engagement proposal supersedes two previous proposals, that more closely focused on the provision of persistent identifier infrastructure for objects accessed through the SRB protocol instead of HTTP. As part of that assessment, PILIN undertook detailed business analysis of likely use cases for persistent infrastructure in e-research, given the fact that SRB already provides appropriate copy functionality and an equivalent of persistent identifiers. The conclusion of this earlier engagement was that the persistent identifiers were most useful to e-researchers outside the curation boundary, especially as a bridge between the representation of objects in HTTP and SRB.

The reingestion of objects from the publication space back into a collaboration space, through resolution of Handles into SRB, was considered out of scope of this engagement. Nonetheless, the previous work mentioned identified a straightforward way of providing this capacity in conjunction with ARCHER work, and PILIN will contribute simple capability to resolve Handles into SRB by the end of the year.

13.1.2 Project Responsibility

13.1.2.1 PILIN

PILIN provides the identifier infrastructure necessary to support the lifecycle of data objects from creation to publication. This includes creating the identifier in the collaboration space (or potentially instrument space). It also means helping manage the identifier as the object crosses the curation boundary—so that PILIN support makes publishing the object through a persistent identifier possible. Moreover, the complexities of managing identifiers across several policy domains require the policy expertise that PILIN has acquired in modeling and analysing identifier usage and best practice.

13.1.2.2 Stakeholder

ARCHER committed one part-time developer for one month to work on the development of integration of infrastructure for creating and managing persistent identifiers, and to undertake knowledge transfer from PILIN staff.

13.1.3 Description of Project Activities and Time Frame

This project was discussed and worked on for a significant part of 2007 but the exact project specifications were not finalised until September.

Activities included:

- Development of the final project proposal
- Provision of policy guidelines and expertise for the management of persistent identifers, and operationality of persistent identifiers spanning multiple policy domains (October-November)
- Business analysis of e-research use of persistent identifiers (May-June)
- Provision of the infrastructure for setting up a Handle server; (September)
- Provision of APIs to allow integration of the Handle server to ARCHER tools; (September)
- Provision of User Interface code for Handle management, to be integrated into ARCHER data management tools (September)
- Assistance to ARROW in assigning persistent identifiers to data objects which already have (incompletely populated) Handles. (November)
- Simple Handle-to-SRB URL resolver, for integration into command-line tools. (November)

13.1.4 Description of Project Outcomes

The project successfully developed work flows and procedures for provisioning with datasets. This work will be extremely useful for a whole range of e-research activities. It will be important for this project to be written up in a paper in 2008 so that it can be shared more fully across the sector.

13.1.5 Evaluative Comments

- **Did the project meet its objectives, describe**

This extension project was critical as an e-research exemplar. The objectives set for the project were fully met at the end of 2007, but more work is required in this critically important field.

- **Did the Stakeholder finish their activity, describe**

Yes, ARCHER and ARROW were able to complete their activity as prescribed by the project proposal. PILIN and ARCHER have moved together to establish the versatility of Handles and their usefulness in the e-research domain. ARCHER has committed only to exploratory work in this engagement, and is satisfied with the outcomes. As a result, ARCHER is well positioned to proceed with implementation work in early 2008.

- **What lessons were learnt for PILIN and future identifier services**

The model of publishing used by PILIN has been refined in light of the identification by ARCHER of a curation boundary: granting access to an object to a group of users needs not count as publishing, and does not call for persistent identification, so long as the group has only a curatorial relation to the object. This was a realisation from the earlier, SRB-specific stage of engagement with ARCHER, which has been confirmed in the current engagement process, and its use of Handles as placeholders rather than as resolvable identifiers at the curation stage.

Technical constraints to do with the mechanics of ingesting data objects were also identified,

as discussed below. The publishing of very large data objects through the library rather than the e-research domain, as required by the project plan, is still an immature field, and there were lessons in this for ARROW in its involvement with the process. These issues are not specific to identifier systems.

- **What unexpected things occurred**

The ARCHER project priorities were significantly altered mid-year, so that publishing objects was no longer in scope for the project. This suspended PILIN (and ARROW) engagement with ARCHER for a couple of months, until the project scope was resolved: ARCHER's engagement is limited to exporting for publication, and ARROW's engagement is to take over responsibility for publication.

The process of ingesting e-research objects into the preferred publication space, ARROW, was onerous because of the size of the data objects, and was anticipated to be infrequent. Therefore there was no interest in the ARROW project automating identifier population as objects with identifiers crossed the curation boundary: the publishing process is to be fully manual. This severely limits the extent to which PILIN machine-to-machine tools (such as APIs) can contribute to the publishing process.

The newly employed senior developer for ARCHER is the former senior developer for PILIN. This has helped the knowledge transfer from PILIN to ARCHER, and has enhanced the collaboration between the two projects. It also ensures that expertise on PILIN deliverables will remain available to ARCHER after the conclusion of the PILIN project.

- **Is there any unfinished business to note**

ARCHER will continue this work into 2008 and the ex PILIN senior developer is contracted through to March 2008.

13.1.6 Summary

ARCHER provided a valuable test bed of identifiers and their role in publishing. This allowed PILIN to refine its notion of publishing and curation, to deal with the complexities of the e-research domain that are not reflected in the conventional repository world. The engagement also helped PILIN better understand and respond to the requirement for incompletely populated identifiers, use of identifiers outside the HTTP protocol, and creation of identifiers outside of registration. ARCHER and ARROW have been able to leverage PILIN work to realise their own workflows more effectively.

13.1.7 Appendix

Appendix 1: Business Requirements Analysis, Use of Persistent Identifiers with SRB

Appendix 2 ARCHER Extension Project proposal

Appendix 3 ARCHER Persistent Identifier workflows

13.1.8 Appendix 1 - Business Requirements Analysis, Use of Persistent Identifiers with SRB

13.1.8.1 Technologies

- Globus Handle Toolkit
- JARGON API
- SRB

13.1.8.2 Assumptions

- e-Research data objects reside on a Collaborative Repository while they are used in active research.
- e-Research data objects move to Archival Repositories when they are no longer used in active research.
- e-Research data objects may move back from Archival to Collaborative Repositories if they are to be used in active research again.
- Collaborative Repositories in the DART/ARCHER use-case uses SRB.
- SRB has its own access tools and API, with underlying protocol distinct from HTTP.
    - An object can be located on SRB through an **SRB URL**, which consists of:
        - A User Name and Password (optional; not relevant for repository use)
        - An SRB Host (specifying one SRB server in a zone or federation)
        - A port number for SRB server connection
        - A Zone (optional; specifying a logical collection of SRB servers and resources, possibly within a federation of zones)
        - A Collection and/or Filename within the Zone (effectively a path)
- SRB paths are logical collections of files and do not indicate physical locations. Zones are likewise logical not physical collections of SRB servers and resources.
    - Therefore SRB paths are persistent locators: a change in physical server location or physical path will not change the SRB path.
        - So a move from collaborative to archival storage repository needs not result in a change of the SRB path
    - SRB paths are semantically meaningful: they have hierarchical directory/collection structure.
    - SRB paths are not persistent identifiers: the logical location (collection/filename) of an object may be changed.
- SRB objects have persistent local identifiers: GUIDs
    - GUIDs may be used instead of SRB paths in some circumstances
    - The GUID remains the same whatever physical or logical path the object is moved to within a zone (SRB federation)
    - The GUID is local to a zone: if an object is moved to a new zone, it receives a new GUID local to a zone (GUIDs include the zone name to prevent GUID clashes)
    - If an SRB object is duplicated rather than moved or replicated, it receives a new GUID
        - This means that an object updated through overwriting will receive a new GUID

- Users interact with SRB through application tools and the Grid. Such interactions generally consist of:
  - Downloading and uploading files.
  - Running files through programs to generate new files. (outside of SRB)
    - The running of files through programs is done through various job scripting protocols (e.g. Nimrod).
  - Many Grid services can process SRB paths.
  - Fewer Grid services can process SRB URLs.
    - The JARGON API for SRB natively processes SRB URLs.
    - If Grid services do not process SRB URLs but only SRB paths, the SRB path will need to be extracted from the SRB URL.
      - This may involve switching context to a different SRB server and zone, providing the user is authorised to access that server and zone.
      - Proof of concept code to do such extraction can be developed
    - Fewer (if any) Grid services can process GUIDs
      - Most do not
      - Proof of concept code to use GUIDs can be developed
      - Aside: GUIDs are not enabled for all SRB servers, however, the GUID can be constructed from the other metadata attributes if it doesn't exist: GUID = ZONE_NAME + ':' + DATA_ID
- SRB access can be authorised and authenticated through a Grid Security Infrastructure (GSI) enabled SRB protocol, or through the standard SRB protocol using username and password.
  - GSI enabled protocols are distinct from web-based protocols.
  - Web services may be granted authorisation through a Grid certificate to initiate requests on the user's behalf, or on the services behalf.
  - (This is a complex issue not fully outlined here.)
- Users interact with the Grid through command line interfaces.
- Users interact with the Grid through Portals, which allow web access to Portlets (e.g. Gridport).
  - Portals are deployed collections of portlet software.
    - Portals may number in the low tens in Australia for the APAC National Grid.
    - Portlet software implementations may not be interoperable.
      - Portals are expected to use instances of the same SRB portlets.
        - So any solutions will need to operate with a limited number of portlets.
  - Portlets allow users to download/upload files, and to schedule jobs to operate on files, through a web browser interface.
  - Individual SRB objects can therefore be accessed through HTTP via a portlet.
    - Portlets *may* allow HTTP hooks into individual SRB objects, i.e. **Portlet+SRB URL**
      - If they do, this hook would be an HTTP URL taking the user to the portlet display of the particular SRB object (collection or file).
      - The feasibility of this requires investigation
- As DART proposes, users will also interact with SRB through Fedora.
  - This will allow repository-like management of data and metadata, which is underdeveloped in SRB

- An SRB object becomes a datastream in a Fedora object
  - The two objects are distinct, since Fedora will also add metadata datastreams
  - (However, the metadata datastream may also be stored in SRB. Requires investigation.)
- The Fedora object is inherently HTTP-accessible, through a **Fedora URL**
  - The SRB object is not inherently HTTP-accessible
  - (Unless the existing Fedora+SRB code base supports REST access to SRB data. Requires investigation.)

13.1.8.3 Requirement

Primary Requirement. e-Research data object need to be cited at various stages of their lifecycle. The citation occurs through an identifier. The identifier should be resolvable within the Grid: i.e. it can be used to get to the actual data object through the Grid. The identifier should resolve correctly wherever the object happens to be stored, and should not need to change if the object changes location. The identifier should be the same at any stage of the lifecycle, including pre-publication.

Secondary Requirement. The identifier should be usable in Web space as well as Grid space: it may appear as a link in a digital object on the web (e.g. a PDF). Clicking the link in Web space should make also it possible for the user to get to the actual data object -- or at least let the user know how to go about it.

Secondary Requirement. The identifier should be rigid rather than fluid. A major motivation for persistent identifiers is the ability to verify an experimental study by rerunning the experiment over given data. The expectation is that once a dataset crosses the "curation boundary" and is published, the referent of the identifier is not updated or augmented, as that would compromise verifiability. The identifier on publication thus acts archivally. (Before publication, it can still act fluidly, as a placeholder for the eventual finalised dataset.)

Secondary Requirement. Identifiers also need to be generated to access workflows processing datasets, not just datasets themselves. A retrieved workflow would typically be fed into a system like Kepler.

So the requirement is for a:

- *persistent* identifier
- which can resolve to objects *in SRB*
- through a resolution service initiated *within the Grid*
  - ***Need to check if this is an actual ARCHER requirement.***
- and which can be actioned usefully *on the Web*

13.1.8.4 Possible Services

There are many ways to address an SRB object, both through the SRB native protocol and through HTTP. The context of use determines the right way to access an object.

There are many services which can interact with SRB objects, including:

- Obtain (through SRB protocol)

- Obtain (through HTTP protocol)
- View Metadata
- Process (through job scheduling protocol)
  - e.g. NIMROD + Gridport + HTTP
  - e.g. CONDOR + Grid (command line) + SRB
  - e.g. KEPLER + Grid (not portalised) + SRB
- Move or Replicate (physical)
- Move (logical)

Given the conditions outlined, the following are possible resolution services, each of which would be used in different contexts.

- Resolve Identifier to Fedora Object URL **[Service 1]**
  - The user views through HTTP the Fedora object containing metadata describing the SRB object, and can access the SRB object itself via the usual mechanism (link, somehow)
    - e.g. The user resolves an identifier found in a PDF, and wants to know what the SRB object is about, who owns it, licensing, and how to get access.
- Resolve Identifier to Portlet+SRB URL **[Service 2]**
  - The user views the SRB object in a Grid Portal through HTTP
    - e.g. The user wants to interact with the SRB object through Gridport (download through HTTP to disk, job schedule through web interface)
    - This may involve a choice of appropriate Portal, if several Portals have been deployed)
- Resolve Identifier to SRB URL **[Service 3]**
  - The user interacts with the SRB object on the Grid or via SRB tools without an HTTP portal
    - e.g. The user accesses the Grid through a command line tool, and wishes to add an SRB object to a job schedule
    - This will likely further involve resolving SRB URL to (appropriate) SRB path
    - For reasons of existing software support, resolution to an SRB path is preferable resolution to SRB GUID, even though the GUID matches persistence requirements better.
    - The requirement for rigid identifiers (experimenters need to know exactly what datasets they are submitting for processing) means it may be a bad idea to have NIMROD interact natively with Handles; better to resolve the Handle to an SRB URL before feeding that URL to NIMROD, so the reseARCHER can inspect the resolution.
- Resolve Identifier to SRB GUID **[Service 4]**
  - The user interacts with the SRB object on the Grid or via SRB tools without an HTTP portal
    - e.g. The user accesses the Grid through a command line tool, and wishes to add an SRB object to a job schedule
    - This will likely further involve resolving SRB URL to GUID (and Zone information, since the GUID is specific to the Zone)
    - This resolution service is an alternative to resolution to SRB URL, given that the GUID is more persistent. If the identifier system maintains a mapping of GUID to

SRB URL, then the two resolution services may be integrated.

- Resolve Identifier + Obtain **[Service 5]**
  - The service has the identifier as input, and returns a datastream of the raw data of the SRB object
    - Such an obtain service will need to deal with Grid/SRB authorisation and authentication, and may be forced to bypass user authorisation
- Resolve Identifier to Fedora Object URL + Obtain to SRB **[Service 6]**
  - As a special case of the above (Resolve Identifier + Obtain), combined with (Resolve Identifier to Fedora Object URL), this service can be used to reingest a data object published to an archival repository (ARROW) back into a collaborative repository (SRB).
    - The service has the identifier as input, and returns a datastream of the raw data of the dataset object *from Fedora*, obtained through an HTTP transport.
    - This dataset object is then ingested into SRB as a separate workflow step.
    - Such an obtain service will need to deal with Grid/SRB authorisation and authentication on the SRB side (ingesting to SRB) through not on the resolve side (datastream from Fedora). As a result, the Ingest to SRB component is fully independent of the resolver system, and does not engage PILIN in Grid/SRB authorisation.

The following value-added services would also be needed for smooth managing of the SRB identifier system.

- Link-rot Checker **[Service 7]**
  - Periodically check that the SRB resource the identifier points to still exists, at the given locator(s).
    - If the locator is an SRB logical path, confirm that the resource has not been moved to a different logical path.
    - If the locator is a GUID, confirm that the GUID remains valid within the zone the resource is located in.
    - If both GUID and SRB logical path are used, confirm that the SRB resource at the given logical path has the given GUID.
- Persistent Citation (non-identifier) **[Service 8]**
  - The infrastructure of Handle can be used to manage GUIDs and SRB URLs as long-term citations; cf. PersistentActionableCitation.
    - In particular, it could be used to map GUIDs to SRB URLs via their common identifier.

13.1.8.5 Use Scenarios

The following use scenario illustrates the use of a persistent identifier to an object on SRB.

Ingest of Raw Data - precondition

1. A data object in molecular biology is gathered through an instrument.
2. The data object is stored on the SRB collaborative repository `srb1`, at the SRB address `srb://mcat1/molecular/object25`
3. The data object has SRB `GUID au/47894`
4. The data object is analysed on the Grid by streaming the object (as identified through its SRB address) into a program. Pre-staging might also be an option.

- eg. `molecular_analyse srb://mcat1/molecular/object25`
- eg. `srbHost=mcat1 ; Scat /molecular/object25 | molecular_analyse`
- eg. `srbHost=mcat1 ; Sget /molecular/object25 ; molecular_analyse < object25`

Registration of ID to SRB

1. ? at Collab Repos or SRB?

Citation

1. The data object is assigned the identifier `SRB_ID:xyzzy`.
2. Jack writes a paper using the data object as a dataset.
3. Jack cites the object in the paper as `SRB_ID:xyzzy`

Resolution within Grid job

1. Jack's Grid portal has a SRB persistent identifier resolution service, `stream_from_ID ___`
2. When Jack calls the resolution service, `stream_from_ID SRB_ID:xyzzy`, the resolution service returns a data stream of the object at `srb://mcat1/molecular/object25` **[Service 5]**
3. When Jack `types stream_from_ID SRB_ID:xyzzy | molecular_analyse`, the data stream of `srb://mcat1/molecular/object25` is fed into the program `molecular_analyse`

Discovery from Citation

1. Jack's paper is published on his institutional repository as a PDF.
2. The PDF hyperlinks the instance of the identifier `SRB_ID:xyzzy` to an online SRB resolution service, `http://resolvesrb.com?id=SRB_ID:xyzzy`
3. Clicking on the hyperlink triggers the online SRB resolution service, which returns the message: "your object is at address srb://mcat1/molecular/object25 . Please arrange access through MCAT1". **[Service 1]**

SRB Link Checker

1. Jack moves the logical path of his object from `srb://mcat1/molecular/object25` to the more descriptive `srb://mcat1/molecular/dioxin/analysis/object25`
2. The GUID `GUID au/47894` still points to the same object, since Jack has not changed zones.
3. A link rot checker for `SRB_ID:xyzzy` determines that `SRB_ID:xyzzy` resolves to the right GUID, but the wrong SRB URL. **[Service 7]**
4. The link rot checker corrects the SRB resolution of `SRB_ID:xyzzy` in the SRB persistent identifier resolver.

From SRB GUID to SRB URL

1. Bjorn has written down the SRB `GUID au/47894` for Jack's object. He does not know about Jack's paper or the identifier `SRB_ID:xyzzy`.
2. Bjorn needs to operate on SRB `GUID au/47894` through a service that requires SRB URL instead of SRB GUID.
3. Bjorn checks the persistent citation service, in case there is a locally available instance of

SRB `GUID au/47894`. **[Service 8]**

4. The local persistent citation service has access to SRB resolution data.
5. The local persistent citation service resolves `GUID au/47894` to `srb://mcat1/molecular/dioxin/analysis/object25` via `SRB_ID:xyzzy`.

From Archive to local

1. The data object moves to an archival repository, which is not on the GRID but is accessed through Fedora. (The backend may still be SRB.)
2. The PDF hyperlinks the instance of the identifier `SRB_ID:xyzzy` to an online SRB resolution service, `http://resolvesrb.com?id=SRB_ID:xyzzy`
3. Clicking on the hyperlink triggers the online SRB resolution service. The object now has an HTTP address instead of an SRB URL. The resolution service returns a metadata page for the object, and the message "Download the data object here".
4. Clicking the download link downloads the data object to local disk. Eventually. **[Service 1]**

From Archive to SRB

1. Tatiana wishes to verify Jack's study. To do that, she needs to get the data object out of the Fedora repository, and back on the GRID.
2. Tatiana feeds the identifier `SRB_ID:xyzzy` to an "ingest to Grid" service.
3. Tatiana gives the "ingest to Grid" service her Grid credentials, and the intended Grid logical path to ingest the data object to.
4. The service resolves `SRB_ID:xyzzy` to the HTTP address to download from Fedora.
5. The service downloads the data object from Fedora.
6. Using Tatiana's credentials, the service uploads the data object to the Grid, in the location specified. **[Service 6]**
7. The "ingest to Grid" service informs the SRB resolution service that the data object `SRB_ID:xyzzy` now has a new SRB URL in a new Zone: `srb://mcat10/molekularnyj/objekt25`

Portal Access

1. Dmitri uses Gridport (portlet) to interact with SRB. Dmitri logs in to Gridport.
2. Dmitri feeds `SRB_ID:xyzzy` from Jack's study to a Gridport ID resolver, in order to access the object.
3. Because Dmitri has already logged in to Gridport, the resolver resolves to the Gridport page displaying the metadata of `SRB_ID:xyzzy` (as `GUID au/47894`). **[Service 2]**
4. From this point Dmitri can inspect the object, filter it, or schedule jobs with it.

Workflow

1. Ulf wants to verify Jack's study. He wants to feed Jack's data and workflow into Kepler. Ulf's Kepler is not portalised, so it cannot use an ID-to-Portal resolver.
2. Ulf obtains Jack's workflow (also through a persistent identifier).
3. Ulf passes `SRB_ID:xyzzy` to an SRB ID-to-SRB URL resolver. **[Service 3]**
4. The SRB resolver returns an SRB URL.
5. Kepler takes the SRB URL and the workflow files as operands, and performs the job.

13.1.9 Appendix 2: ARCHER proposal for stakeholder engagement with PILIN

2007-08-31
Andrew Treloar, Nick Nicholas

ARCHER is interested in exploring procedures and workflows for the provisioning of persistent identifiers for datasets, both before and after those datasets cross the curation boundary into a publication repository.

13.1.9.0.1 Scope:

The particular use case ARCHER wishes to explore is:

- creation of identifiers in the research collaboration space
- managing the transition of the identified objects from the collaboration space into the publication space

There are other use cases which are not in scope of the proposed engagement:

- Moving an identified object from a private reseARCHER space to a collaboration space
- Moving an identified object from a publication space back into a collaboration space

The creation of identifiers may precede publishing the object, but this is to occur on an as-needs basis. There is no expectation that of blanket generation of persistent identifiers for objects in the collaboration space.

There is no expectation that persistent identifiers will be used to manage objects in the collaboration space in general. The identifiers created in the collaboration space may be resolvable (e.g. to SRB URLs), but at this stage such resolution will not be critical to the collaboration processes: the persistent identifiers will be used only for publishing activities, outside the curation boundary. For example, there is no requirement of command-line resolution of persistent identifiers, or integration of persistent identifiers with job scheduling scripts.

There are three distinct contexts that identifiers in this process may belong to:

- Identifiers under the control of the publication space (e.g. ARROW hdl:1959.1/)
- Identifiers under the control of the collaborative space (e.g. ARCHER hdl:/102.100.ARCHER)
- Identifiers under the control of the participating reseARCHERs (e.g. MIT hdl:1721/ , if an ARCHER reseARCHER later relocates their objects to MIT)

Working through the governance alternatives for such identifiers is an important output of the engagement. Therefore the identifiers generated may not be constrained to one namespace.

13.1.9.1 Deliverables and Time Frame

1. **The extension project will commence in September 2007 and be complete before December 2007.**
2. **The PILIN project team will provide support for ARCHER so that the following deliverables can be achieved by ARCHER**

1. Support for updating identifiers of objects moving from collaboration space (ARCHER) to

publishing space (ARROW):

- Support for ARROW workflows for updating the resolution of ARCHER-created handles to point to ARROW instances of ARCHER objects, as ARROW ingests the objects.
- Automation may mean some integration of Handle API tools with ingest scripts

PILIN will supply developer time to assist ARCHER with this work

2. A report detailing the identifier policy issues involved in the desired workflows, and the options ARCHER might take on identifier governance. This report shall be integrated into the PILIN project outputs, and released publicly. The issues include:

- How to ensure an identifier is persistent in reference after its referent is no longer under one s control (aliasing options, shared identifier management options)
- Timing and granularity of identifier assignment
- Should identifiers be withdrawn and recycled
- Should two identifier contexts share a common namespace, and if so how (e.g. ARROW sequence numbers vs. PILIN timestamps under the same Handle namespace; cf. PILIN Curation & Hosting models document)

PILIN will provide business analyst and information manager time to assist with this work.

3. Use cases corresponding to the various options outlined above, describing what possibilities the deliverables support.

PILIN will provide business analyst time for this work

4. A prototype **investigation** of integration of identifier generation tools into a collaborative space. The collaborative space shall MAY be crystallography at Monash. The integration shall be a proof of concept, with no expectation of production-ready code. The code should be publicly accessible as a demonstrator.

The integration may be either with asset management tools developed by the crystallographers at Monash (contact: Mark Bate), or with the tools in place in DART/ARCHER.

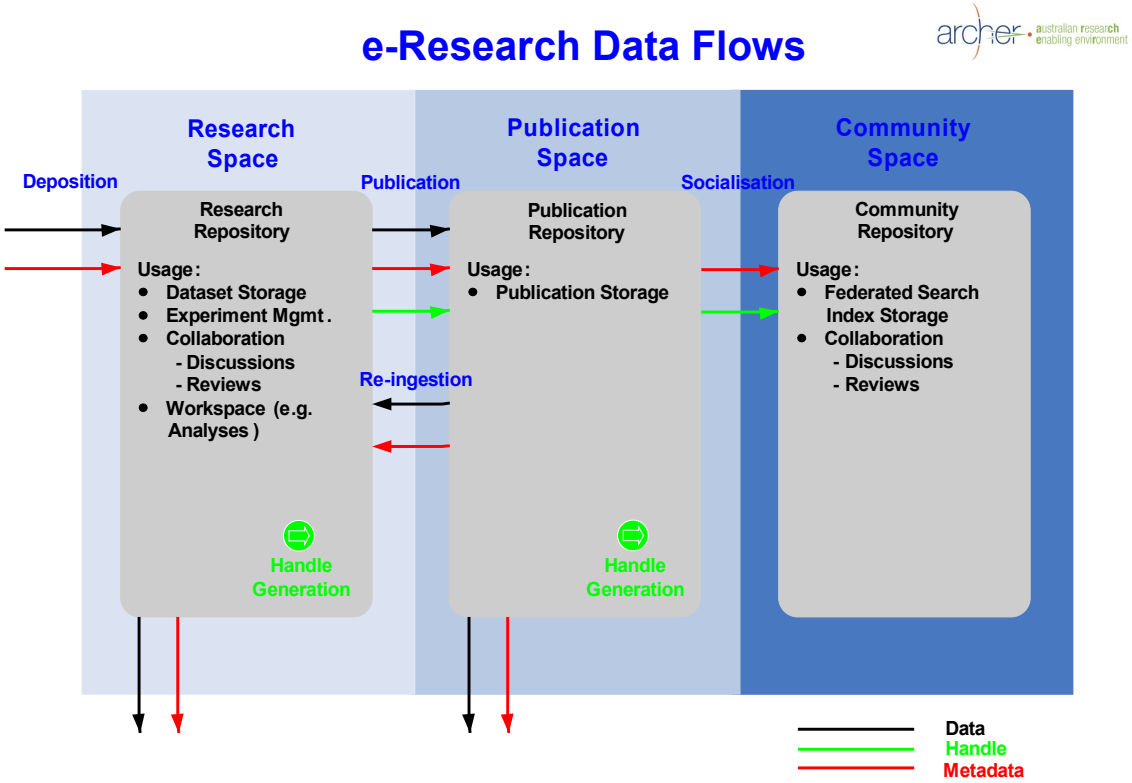
The generation task is of lower priority than the publishing integration task under (1); it may be satisfied by manual creation of identifiers through an unintegrated tool PILIN makes available to the crystallographers.

PILIN will provide developer time to assist ARCHER with this work should it become a priority

13.1.9.2 Summary

It is estimated that PILIN will provide a maximum of 10 days developer time and 10 days of business analyst time (as well as some technical advice from other team members for this extension project for a total estimated cost of $15,000

13.1.10 Appendix 3: ARCHER Persistent Identifier workflows



The e-Research space can be seen as being composed of three distinct sub-spaces: research; publication; and community.

In the *Research Space*, researchers deposit datasets from sources such as scientific instruments, CDs, or desktops. This repository is also used for managing experiments, collaborations between researchers, and as general workspace. Once research content is ready for public exposure, it is transferred to the *Publication Space* through the publication process. In the *Community Space*, research communities can easily discover, discuss, and review research content relevant to their field.

The diagram above shows the flows of data, handles, and metadata into, out-of, and between these spaces.

Handles are useful in the e-Research space in the following ways:

- They may be generated within the research or publication repositories to enable referencing within their associated spaces.
- During the publication process, research content and its metadata are normally transferred to a publication repository, however, some datasets may be quite large and it may be more suitable to transfer the dataset s handle instead.
- During the socialisation process, services in the *Community Space* only require references to research content (handles) and the associated metadata.

13.2 State Library NSW
http://www.sl.nsw.gov.au/

pilin persistent identifier linking infrastructure

web: http://resolver.net.au/hdl/102.100.272/TBA

email: policy@pilin.net.au

Version History

Version	Date	Status & changes	Expression identifiers
V0.1	2007-10-22	DRAFT: Initial draft for internal discussion.	PILIN/???? hdl:102.100.272/TBA
			PILIN/TBA hdl:102.100.272/TBA

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

Stakeholder	Name of Project	Resources used in terms of time estimate of $	$ allocated out to stakeholder
SLNSW	**PILIN Persistent identifiers –Digital Asset Management (DAM)**	**$8K** **Made up of Information manager and Business Analyst time**	**0**

13.2.1 Description of Project and Objectives

Looking at using Handle base for identifiers within Digital Asset Management system including levels of aggregation, points of assigning Handle, multi point object and identifier, semantics, relationships and different format versions.

Original submission from SLNSW attached as appendix A.

13.2.2 Project Responsibility

13.2.2.1 PILIN

PILIN to supply advice and support for project scoping and implementation in form of business analyst and information manager time.

13.2.2.2 Stakeholder

Contacts at SLNSW have been:

Katie Wilson,

Simon Handfield

Suzanne Moir

Lynne Billington

Richard Neville

Shirley Walker

13.2.3 Description of Project Activities and Time Frame

A number of meetings were held.

In **January**, the following issues were discussed

- Currently don t have persistence
- Don t have expertise to design own system
- Designers want to be told what to do, but we don t know what to tell them.
- Have spoken to National Library of Australia and National Lib of NZ
- Want to use a standard but there is none
- What happens to legacy identifiers?

DAMS – Digital Asset Management System was put forward a use case for the State Library of New South Wales. SLNSW will complete a use case template when Nick circulates the use case exemplars.

In **May**, there was further discussion of SLNSW requirements. It was agreed that PILIN would work with SLNSW developers, and that PILIN deliverables to them are source code plus documentation of PILIN interface.

In **July,** there was a teleconference with PILIN team, including Business Analyst, Senior Developer and Information Manager, SLNSW Team (Katie Wilson, Simon Handfield, Lynne Billington) and Sunith Miranda, Project Manger, Media Equation – implementation firm for SLNSW. They reviewed PILIN: JAHDL 1.0 [JAVA API for Handle] and had the following response:

> We ve reviewed the possibility of integrating this with lookat.me we can pass xml requests/files as parameters to their command line tool PilinHandleMgmtTool
> Possible functions include CRUD [Create, Resolve, Update, Delete] with batch operations.
> Create functionalities could be invoked during digital object validation, ingestion or via a control panel in lookat.me to initiate selected CRUD operations.

A further meeting was held in **August** in Sydney. It was decided that before any more progress could be made, SLNSW would have to decide on an 'information model'. It was agreed that draft PILIN policy documents would be distributed to SLNSW and that SLNSW would make use of Nick Nicholas to further the SLNSW information model.

13.2.4 Description of Project Outcomes

The project is still in operation. PILIN's contribution will cease, at the end of 2007 but SLNSW are likely to tap into the National Persistent Identifier Service if that is established. It is anticipated that the SLNSW project will conclude during 2008.

13.2.5 Evaluative Comments

- **Did the project meet its objectives, describe**

Engagement with the PILIN team highlighted a number of policy decisions that have to be made by the State Library of New South Wales. The PILIN team contribution has helped focus the SLNSW work in terms of defining the information model.

- **Did the project finish, describe**

The policy and information modelling required by State Library of New South Wales use of persistent identifiers has been advanced through engagement with PILIN. However, as a result of installation of new technolgoy platform, the implementation phase of the engagement project has been delayed to the point where there are no concrete outputs of the engagement to report.

- **What lessons were learnt for future identifier services**

The extension project focussed PILIN attention on the need for documenting expertise on persistent identifiers in policy guidelines so that they are available for wider consultation.

- **What unexpected things occurred**

Katie Wilson, the main contact for PILIN at SLNSW, left the library at the end of September. As the library is installing a new technology platform, this has taken precedence.

- **Is there any unfinished business to note**

Yes - the project is still to complete.

13.2.6 Summary

The project was valuable for PILIN in identifying the need for a sound underpinning information model. The PILIN input was valuable for SLNSW in that the policy guidelines and software toolkits produced by PILIN will underpin the SLNSW implementation. At the same time the PILIN business analyst and the PILIN information manager helped shape the DAM project at SLNSW

13.2.7 Appendix

Appendix 1: PILIN persistent identifiers Proposal, State Library of New South Wales

13.2.8 Appendix 1 - PILIN persistent identifiers Proposal

Document version 1.0 – Release

Date 3 May 2007

13.2.8.1 Document control

13.2.8.1.1 Document details

Attribute	Value
Project Name	PILIN persistent identifiers
Project ID	H2
Document Title	Proposal
Document Version	1.0
Version Date	May 03, 2007
Document Status	Release
Author/s	Katie Wilson, Susanne Moir, Simon Handfield

13.2.8.1.2 Document history

Version history	Date published	Change History
0.1	20 April 07	Working Draft
1.0	3 May 2007	Release

13.2.8.2 Proposal for PILIN extension project

The State Library of New South Wales has a unique history with origins dating back to 1826 and a collection of over five million items. The Library is a premier information source for the people of New South Wales and beyond. It is governed by a Library Council with an Executive responsible for day-to-day operations. The State Library collection is a nationally significant resource.

The Library is undertaking a progressive program to deliver client focused e-services. It is developing a new sustainable technology platform to support an extensive digitisation program and expand the library's e-service delivery.

13.2.8.3 atmitchell.com project

The State Library of NSW (SLNSW) has been digitising, managing and publishing content, including web content, since 1999 when a special unit was formed to create and deliver resources in digital form. By 2001 the Library was producing 2,050 digital images per annum. In 2006 the number of digital masters scanned in a year had increased to 24,087. The images are accessible on the Library's website through exhibition-style web pages which provide contextual and curatorial information; as well as via links from records in the Library's online catalogues.

With the new technology platform the State Library's atmitchell.com project aims to increase digitisation of unique archival material from the library's collection over the next three years. A new Digital Asset Management System (DAMS) and Archival Collection Management System (ACMS) are currently being implemented, together with a Web Content Management System (WCMS) and Business Process Management software (BPM).

13.2.8.4 Persistent identifiers for digital objects

As the Library moves forward in the development of the digital repository it has recognised the need to put in place a robust scheme of unique and persistent identifiers for its digital objects to maintain their accessibility over time.

13.2.8.5 PILIN aims

The PILIN Project aims to support the adoption and use of persistent identifiers and shared persistent identifier management services by the project stakeholders. It plans to develop a sustainable, shared identifier management infrastructure that enables persistence of identifiers and associated services over archival lengths of time.

PILIN is building identifier management infrastructure based on a technology (Handle) that is now under development through the auspices of CNRI to underpin sustainable global identifier infrastructure. The project will develop use cases describing community requirements for identifiers and business process analysis relating to these use cases.

13.2.8.6 Collaboration proposal

The State Library of NSW wishes to propose collaboration on an extension project with the PILIN project team on the atmitchell.com digitisation project as a use case.

13.2.8.7 Use cases

In the proposed test case the Library will provide a new environment to test the generation of persistent identifiers for digital objects that are stored in the Digital Asset Management system (lookat.me), and their integration with the PILIN tools via web services.

Digital objects are linked to the library's website in online exhibitions, journeys, and catalogues such as the Archival Collection Management System, and that for published materials (Millennium). In the proposed extension project the Library and PILIN will be able to test how these key systems integrate with the Handle Persistent Identifier infrastructure.

The Library also wishes to address and review the most appropriate level of granularity and timing for assigning persistent identifiers, addressing questions such as:

- Should persistent identifiers be assigned to all digital object derivatives?
- At which collection level should identifiers be assigned to digital objects?
- Should we allow meaning to be embedded in persistent identifiers?
- Does a persistent identifier have a one-to-one relationship with a single object?
- Can one persistent identifier point to more than one object?
- Can more than one persistent identifier point to the same object?
- At what point in an object s life is it assigned a persistent identifier?

13.2.8.8 SLNSW contribution

State Library of NSW will contribute a database and software that is using a different technology, with defined digitisation processes, and State Library development staff.

13.2.8.9 What we need from PILIN

Advice and expertise from PILIN developers to integrate the State Library's Digital Asset Management system with the persistent identifier tools that PILIN is developing.

13.2.8.10 Timeline

July/August 2007

13.3 The Learning Federation (TLF) Schools Sector

http://www.thelearningfederation.edu.au/

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
V0.1	Yyyy-mm-dd	DRAFT: Initial draft for stakeholder discussion.	PILIN/???? hdl:102.100.272/TBA
V0.2	Yyyy-mm-dd	Changed fonts	PILIN/TBA hdl:102.100.272/TBA
V1.0	Yyyy-mm-dd	RELEASE: Initial release to public	PILIN/TBA hdl:102.100.272/TBA

To cite the *latest* version of this work use http://resolver.net.au/hdl/102.100.272/TBA

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

Stakeholder	Name of Project	Resources used in terms of time estimate of $	$ allocated out to stakeholder
TLF - The Le@rning Federation	Developing a Handle based appropriate copy service.	Up to $8k of PILIN time and $10K allocation to FRED project in kind) PILIN Business Analyst and developer time as well as software – up to 12 days. PILIN resource to FRED so that it could participate in this Extension Project until end of 2007	Nil

13.3.1 Description of Project and Objectives

The TLF distributes its content to various jurisdiction-based portals around the country. Each copy of a distributed TLF resource has its own access infrastructure and URL, and copies distributed to one jurisdiction are not available from another. The TLF has a central access point for its content (the TLF EC), but does not wish this central point to take over from the established jurisdiction distribution network. The multiple distributed copies of TLF resources make it impractical to hyperlink to TLF content from syllabi, library catalogues, or e-mail: a link to a resource in any one jurisdiction will be inaccessible from members of other jurisdictions, and the TLF does not want the maintenance burden of separate links to its content from all its stakeholders, if possible. This means that there are limits on sharing, reuse, and leveraging of TLF content.

To address this issue, The TLF Phase Three Plan introduced the concept of a central registry and metadata repository. In light of this, a pilot appropriate copy delivery mechanism was proposed to TLF, using both PILIN and FRED resources. TLF content will have an "instance registry" of all available copies of TLF content, located centrally. The FRED project provides Appropriate Copy resolution of the instances retrieved from the instance registry, so that a request for TLF content will be routed to the instance registry, and from there to deliver the instance appropriate to the requester (typically, the copy from the local jurisdiction). Appropriate Copy resolution is through the OpenURL protocol. The FRED project contributes the OpenURL toolkit created within the project, and developer advise on installation. FRED also contributes BA time for user requirement analysis, documentation of the OpenURL service expression interface; documentation of the proposed architecture as a service usage model; and feedback on the TLF's own requirements analysis, architecture, and development.

13.3.2 Project Responsibility

13.3.2.1 PILIN

TLF content will have peristent Handle identifiers not specific to any one jurisdiction copy, and the Handle system acts as the instance registry of all available copies of TLF content, resolvable to jursidiction content via an appropriate copy selection service. The instance registry is to be set up as an identifier system with PILIN assistance and PILIN tools. PILIN Business Analyst and software developers worked closely with TLF on this project.

13.3.2.2 Stakeholder

Preety Agarwal, Technical Project Manager and a software developer provided TLF input into the project and attended meetings with PILIN and came to PILIN events.

13.3.3 Description of Project Activities and Time Frame

The project commenced in early 2007 and completed in late 2007. Numerous meetings and exchange of ideas occurred. This project was characterised by both PILIN and FRED involvement. Activities included:

- demonstrations and global discussion of the proposed solution with TLF (March) and with TLF stakeholders (Learning Federation Consultative Committee: April)
- Provision of the infrastructure for setting up a Handle server; (May)
- Provision of APIs to allow integration of the Handle server to TLF asset management systems; (May)
- development team advice on deployment; (May-June)
- integration of APIs with FRED Appropriate Copy toolkit, to be delivered together to TLF (July)
- BA advice on strategies to ensure prompt updating of the identfier records. (It was determined that the jurisdictions will need update access to a TLF instance registry). (September)
- BA advice on use cases for identifiers in the context of appropriate copy delivery. (September-October)
- Review of TLF requirements and specification (September-November)
- Demonstration of TLF prototype at PILIN stakeholder day on 21 November by TLF staff

13.3.4 Description of Project Outcomes

The TLF have built an appropriate copy prototype and will deploy as a sustainable service in 2008.

13.3.5 Evaluative Comments

- **Did the project meet its objectives, describe**

While it took more elapsed time than was first envisaged, the project met its objectives in that the TLF have identified a solution for a real issue and are in a position to deploy in 2008.

- **Did the Stakeholder finish their activity, describe**

National deployment will occur in 2008, but the TLF identified a solution and built a demonstrator based on PILIN/FRED toolkits during 2007.

- **What lessons were learnt for PILIN and future identifier services**

The extension project provided a concrete use case for multiple resolution. This is a requirement common to several domains (intrinsic to content delivery in federations and in decentralised content delivery in general), but which is not supported directly by all identifier schemes. The use case provided a focus point for modelling multiple resolution for both the PILIN and the FRED projects. It also provided an opportunity to extend the OpenURL protocol to the e-learning domain, where it has not previously been extended.

The ability for multiple stakeholders to share management of an identifier scheme (the "Federated" model of identifier management defined in PILIN guidelines) is necessary for a loose federation such as that of the TLF content delivery points. This provides a policy challenge in coordinating multiple parties from disparate institutions with write access to the same identifier system --- a scenario not anticipated in the normal, institution-bound deployment of identifier management systems. (The move to federated service management is also a challenge to conventional thinking on OpenURL.)

There is also a policy challenge in encouraging members of a loose federation not to cite resources through their local URLs but through a central appropriate copy service instead. The benefits for citing through an external third party are apparent to the content provider (the TLF) and to collaborators between jurisdictions; they are not as apparent to the content delivery point (the jurisdictions), who now have to rely on an external service for delivery of their locally-hosted content. An actionable citation (reverse lookup) service was identified as a possible solution to this issue.

- **What unexpected things occurred**

The TLF initially had difficulty obtaining the services of a business analyst, which delayed development from the proposed schedule (August - October). But work was completed during the proposed timeframe.

Also since the toolkits from PILIN and FRED were delivered to TLF late, the initially scheduled development opportunity window was assigned to another TLF project.

- **Is there any unfinished business to note**

The TLF will continue to deploy PIs in its future work and will no doubt be a user of any National PI Utility service.

13.3.6 Summary

The TLF extension project helped focus PILIN and FRED thinking on a problem acutely felt by

federations as content delivery points. The PILIN/FRED solution has helped resolve this problem for the TLF elegantly, with efficiencies of cost and resources. It has also contributed to a better understanding in the user community of persistent identifiers not specific to a single institutional domain.

13.3.7 Appendix

Appendix 1: Business Requirements Analysis, TLF

Appendix 2: Proposed Solution, TLF

Appendix 3: PILIN & TLF collaboration Project Proposal

13.3.8 Appendix 1 - Business Requirements Analysis, TLF

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
V0.1	Yyyy-mm-dd	DRAFT: Initial draft for stakeholder discussion.	PILIN/???? hdl:102.100.272/TBA
V0.2	Yyyy-mm-dd	Changed fonts	PILIN/TBA hdl:102.100.272/TBA
V1.0	Yyyy-mm-dd	RELEASE: Initial release to public	PILIN/TBA hdl:102.100.272/TBA

To cite the *latest* version of this work use http://resolver.net.au/hdl/102.100.272/TBA

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

13.3.8.1 Background

- TLF content is duplicated to jurisdiction repositories.
- In some cases, TLF content is further duplicated from jurisdiction repositories to repositories in school districts or individual schools.
- TLF content is also accessible on the TLF EC by arrangement. (This is not the default distribution path.)
- TLF content is primarily accessed online, through the web (i.e. through hyperlinks).
- The URL for accessing TLF content on jurisdiction and school repositories is decided by local repository managers, not by TLF.
- Access to content individual repositories is restricted. For example, a jurisdictional repository is only accessible to a teacher in that jurisdiction.

13.3.8.2 Requirement

The TLF would like to provide links for accessing its content independently of where the content is stored (i.e. whether it is in jurisdiction repositories, school repositories or the EC). These links would be provided in

- Metadata provided as part of TLF content
  - "Catalogue Entry" for the content being described
  - "Relation" for related content
- SCIS records, which will be copied to school library catalogues.

Additionally, users of TLF content would like to provide links for accessing content independently of where the content is stored. These links would be provided in

- Emails between teachers who may not have access to the same jurisdiction/school repositories.

13.3.8.3 Issues

Currently links are only possible to specific locations -- that is, to specific repositories. The TLF cannot currently provide a single link which will resolve to a copy in a repository which any given user has access to. Therefore the TLF cannot provide a single link fulfilling the requirement stated above.

13.3.8.4 Possible solution: individualised metadata and SCIS records

- The TLF manages distinct links for each jurisdiction or school repository that contains its content.
- The TLF would need to provide distinct metadata and distinct SCIS record with customised links to each jurisdiction or school repository.
- Emailed links refer to individual repositories and can only be used by users who access to that repository.

13.3.8.5 Possible solution: Appropriate location service

The TLF provides a single link in the metadata and SCIS records which has the following functionality:

- When a user clicks a link, it should resolve to the copy of the learning object in a repository which that user can access.
- If the user normally accesses TLF content through a Jurisdiction repository, the link should resolve to the copy on the Jurisdiction repository.
- If the user normally accesses a particular TLF object through a local school repository, then when a user clicks on any link the TLF provides, it should resolve to the copy on that school repository.

This solution requires that a service mapping TLF content identifiers to content locations must be maintained, for each content location in Australia (jurisdiction, school district, or school).

This service would reduce the TLF's maintenance overhead.

- The TLF does not need to provide distinct metadata and SCIS records for each jurisdiction or school repository.
- The TLF does not need to update its metadata and SCIS whenever the location of an object in a repository changes.

This type of link that resolves to a copy of a resource most appropriate to the user, is referred to as an *appropriate location* link. This is a type of *appropriate copy* link.

13.3.8.6 Scenarios

The following are scenarios where such links are required.

13.3.8.6.1 Appropriate Location resolution from metadata record: Link to Content Object

Lois, a teacher at Cairns High, is searching Learning Place, the Queensland jurisdictional server, for learning objects on natural disasters. As a result of her search query, she sees a list of matching objects with short descriptions. The object which strikes her as relevant is TLF123, a learning object that was created at TLF. The list provides links to access the actual learning objects. (Access can mean viewing the object on a browser, or downloading it.) Lois clicks on the "View this" link to view the actual learning object.

The TLF EC has a copy of TLF123. Learning Place has a copy of TLF123. Cairns High repository also has a copy of TLF123. The Learning Place and Cairns High objects are duplicates of the TLF EC object, apart from some minor differences in presentation and branding.

When Lois clicks the link to "View this" on the Learning Place search page, the copy of TLF123 she gets access to is not the copy on Learning Place or TLF EC, but the copy at Cairns High. That is because the appropriate location service has determined that the copy at Cairns High is the copy most appropriate for her.

13.3.8.6.2 Appropriate Location resolution from metadata record: Link to Related Object

Shane, a student at Cairns High, is looking at a description of a learning object on volcanoes. The description is the front page of the learning object he has been directed to by his teacher Lois. He is accessing the learning object through Learning Place, the Queensland jurisdictional server. The learning object was created at TLF, with identifier TLF123.

Narelle, a student at Tweed Heads High, is looking at the same learning object on volcanoes, accessing it through TALE, the NSW jurisdictional server.

Aloysius, a student at St Alfonso's, Hobart, is accessing the same learning object on volcanoes through the TLF EC.

The Learning Place and TALE objects are duplicates of the TLF EC object, so their metadata is identical. TLF have embedded "see also" links in the metadata of TLF123 ("Relation" field) to other related objects; for instance, TLF124, a learning object on Sicily. This "Relation" field is used to create a hyperlink to the object in the description Shane, Narelle, and Aloysius are all looking at.

Learning Place, Cairns High, and TALE both have a copy of TLF124. When the student clicks on these links (which are the same at Learning Place, TALE, and TLF EC), they get metadata about the learning objects, including an abstract, and the option to access the *local* full version where available. If Shane clicks the link, he gets access to the copy of TLF124 at Cairns High. If Narelle clicks the link, she gets access to the copy at TALE.

13.3.8.6.3 Appropriate Location resolution from library catalogue

Cairns High and Tweed Heads High add SCIS records for TLF124 to their catalogues. The same record prepared by the TLF is ingested by both high school library catalogues. The SCIS record has a link to the object referred to. When a student clicks on the link at Cairns High, they are taken to the Cairns High-specific copy of TLF124 (either at Learning Place or in the Cairns High local repository). Similarly, when a student clicks on the link at Tweed Heads High, they are taken to the Tweed Heads High-specific copy (either at TALE or at the Tweed Heads High local repository).

13.3.8.6.4 Appropriate Location resolution from email

Lois, a teacher at Cairns High, has found another object of interest on Learning Place: TLF127, a learning object on the tarantella. She lets Bill, a teacher at Tweed Heads High, know about TLF127 by mailing him a link. Since Bill is at Tweed Heads, his appropriate copy of TLF127 is at TALE. When Bill clicks the link in the email, it resolves to a TALE copy of TLF127, if available. If not, Bill sees metadata about TLF127.

13.3.8.6.5 Appropriate Copy resolution from teacher constructed content

After the passing of Steve Irwin, Bill, a teacher at Tweed Heads High, decides to put together a list of relevant learning objects, along with some commentary on how each learning object can be used in the classroom. (This is a new, *aggregate* digital object, consisting of component learning objects, and independently contributed metadata.) Bill uses TLF content for his learning objects in the list. TLF127 is one of the included learning objects. Bill sends his list + commentary as a single digital object to Lois at Cairns High. When Lois attempts to access TLF127 on Bill's list in Cairns, she accesses the Learning Place copy of the resource. When Bill in Tweed Heads attempts to access TLF127 on the same list, he accesses the TALE copy of the resource.

13.3.8.6.6 Changing Locations of Resources (Persistence)

Learning Place decides to split its storage of TLF content across two geographically separate servers.

Learning Place gathers the new URLs of its objects, and updates whatever mechanism is used to resolve "appropriate location" links correctly in its repositories. The TLF content objects at Learning Place are not updated. The links to other TLF content ("Related Objects") within TLF content objects held at Learning Place continue to resolve to the right copies.

SCIS records distributed throughout Queensland link to TLF content. The SCIS records are not updated. Any SCIS records which point to TLF content held at Learning Place continue to resolve to the right copies. Access to TLF content is not disrupted in Queensland.

13.3.8.7 Lower Priority Narratives

The following are value-added services enabled through persistent identifiers that are also relevant to TLF. They do not directly involve Appropriate Copy services, and the TLF has identified that they are lower priority.

13.3.8.7.1 Sharing Third Party Metadata

Bill at Tweed Heads High provides a ranking for TLF127 through the TALE portal. Because TALE uses the TLF identifier to refer to the resource, his ranking can be fed back to TLF easily, and made available to consumers of the TLF through other jurisdictions, such as Lois at Cairns High.

13.3.8.7.2 Rights Management

This is addressed through TLF's CRISP initiative; the terms under which content is distributed suggest that the TLF identifier infrastructure is already addressing this requirement.

13.3.8.7.3 Appropriately licensed copy

The licensing arrangements of individual jurisdictions will also determine which is the most appropriate copy (if a user has access to more than one). If a jurisdiction has a copy of the object but the user is not authorised to use it, that means the user shall not be redirected to that copy.

This is addressed through TLF's CRISP initiative; the terms under which content is distributed suggest that the TLF identifier infrastructure is already addressing this requirement.

13.3.8.7.4 Appropriate version

A user may be directed to different versions (customisations) of the object, depending on their profile; e.g. a Form 3 version vs. a Form 6 version. Managing such an appropriate *version* service is contingent on a relationship service with an appropriate metadata ontology.

13.3.9 Appendix 2 - Proposed Solution, TLF

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
V0.1	Yyyy-mm-dd	DRAFT: Initial draft for stakeholder discussion.	PILIN/???? hdl:102.100.272/TBA
V0.2	Yyyy-mm-dd	Changed fonts	PILIN/TBA hdl:102.100.272/TBA
V1.0	Yyyy-mm-dd	RELEASE: Initial release to public	PILIN/TBA hdl:102.100.272/TBA

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

13.3.9.1 Assumptions

- There is a service providing appropriate copy resolution of TLF identifiers to Jurisdiction Copies of identifiers. This may be hosted centrally (TLF) or locally (at the jurisdictions).
- The Jurisdictions are responsible for keeping this service up to date.
- As a simplifying assumption, only TLF content is considered in scope for TLF identifiers, and the jurisdictions provide appropriate copy resolution only for TLF identifiers. (Expanding the scope of the service is not a technical challenge but a policy challenge).
- Customisation and leveraging of content is not addressed here: only sharing. Customisation and leveraging presuppose a relationship service.

13.3.9.2 Appropriate Copy resolution in metadata record (by Location)

Shane, a student at Cairns High, is looking at a learning object on volcanoes. He is accessing the learning object through Learning Place, the Queensland jurisdictional server. The learning object was created at TLF, with identifier TLF123.

Narelle, a student at Noosa High, is looking at the same learning object on volcanoes, accessing it through TALE, the NSW jurisdictional server.

Aloysius, a student at St Alfonso's, Hobart, is accessing the same learning object on volcanoes through the TLF EC.

The Learning Place and TALE objects are duplicates of the TLF EC object, apart from some minor differences in presentation and branding. In particular, their metadata datastreams are identical. TLF have embedded "see also" links in the metadata of TLF123 ("Relation" field) to other related objects; for instance, TLF124, a learning object on Sicily. Objects are linked through identifiers, not hyperlinks: the links need to work wherever the student is located, so long as the jurisdictional server knows about the object.

Learning Place and TALE both have a copy of TLF124. When the student clicks on these links (which are the same at Learning Place, TALE, and TLF EC), they are meant to get metadata about the

learning objects, including an abstract, and the option to access the *local* full version where available. Access can consist of viewing the object on a browser, or downloading it.

Notes

The appropriate copy service may reside in the jurisdictions or centrally.

If it resides in the jurisdictions, the server name will need to be populated at runtime.

As an extension: the licensing arrangements of individual jurisdictions will also determine which is the most appropriate copy (if a user has access to more than one). If a jurisdiction has a copy of the object but the user is not authorised to use it, that means the user shall not be redirected to that copy.

As an extension: a user may be directed to different versions (customisations) of the object, depending on their profile; e.g. a Form 3 version vs. a Form 6 version. Managing such an appropriate *version* service is contingent on a relationship service with an appropriate metadata ontology.

13.3.9.3 Appropriate Copy resolution of external reference (catalogue)

Cairns High and Noosa High add SCIS records for TLF124 to their catalogues. The SCIS record has a link to the object referred to in the form of an identifier, for the same reason as above: the SCIS record is populated at TLF, and an identical record is disseminated to all the jurisdictions and schools.

Notes

- The appropriate copy service may reside in the jurisdictions or centrally.
- If it resides in the jurisdictions, the server name will need to be populated on ingestion.

13.3.9.4 Appropriate Copy resolution of external reference (email)

Lois, a teacher at Cairns High, has found another object of interest on Learning Place: TLF127, a learning object on the tarantella. She lets Bill, a teacher at Noosa High, know about TLF127 by mailing him a link. The link should resolve to a TALE copy of TLF127, if available. If not, the user sees metadata about TLF127.

Notes

- The appropriate copy service may reside in the jurisdictions or centrally.
- If it resides centrally, a single email link can be generated; the resolver will work out the correct jurisdiction to invoke when the service is activated
- If it resides in the jurisdictions, the server name will need to be part of the email link. There will need to be a different email link for each jurisdiction, though the query parameters will be the same.
- A central redirection service could be orchestrated with local appropriate copy services, without needing to deploy a full appropriate copy service centrally

13.3.9.5 Appropriate Copy resolution of aggregate objects

After the passing of Steve Irwin, Bill, a teacher at Noosa High, decides to put together a list of thematically relevant learning objects, along with some commentary on how each learning object can be used. This is a new, aggregate digital object, consisting of component learning objects, and independently contributed metadata (the commentary). Bill uses TLF content for his component learning objects. The components are accessed through identifiers, not locators or embdedded content. Bill sends his aggregate object to Lois at Cairns High. Because identifiers are used, an attempt to

access a component in Cairns will take the user to the Learning Place copy of the resource, while an attempt to access a component in Noosa will take the user to the TALE copy of the resource.

Notes

- As above
- The links to objects should take the same form as the links to objects in metadata seen above; there is no real functional distinction between the two cases, so they should share infrastructure for dealing with them.

13.3.9.6 Changing Locations of Resources (Persistence)

Learning Place decides to split its storage of TLF content across two geographically separate servers. Learning place updates its appropriate copy service to the new URLs for TLF content. TLF content is only ever cited or referred to by identifier, rather than URL, within Learning Place (including the "Related" metadata fields of content). Similarly thwe SCIS records distributed throughout Queensland link to TLF through Appropriate Copy Service Calls as their URLs, rather than static locators. Access to TLF content is not disrupted in Queensland.

12.2.10 Appendix 2 - DLIN & TLF collaboration

THE le@rning FEDERATION

schools online curriculum content initiative

Curriculum CORPORATION

Approvals

Name/Organisation/Position	Signature	Date
Preety Agarwal Technical Project Manager - TLF **(Author)**		29th March 2007
Nick Weideman Chief Technology Officer - CC **(Authoriser)**		29th March 2007

Management Summary

The Le@rning Federation is an initiative delivered on behalf of the Australian Education Systems Officials Committee (AESOC) by Curriculum Corporation. The Le@rning Federation Phase Three Plan involves development of a central registry and metadata repository which would enable Australian and New Zealand school educational jurisdictions to provide enhanced discovery mechanisms for their teachers. This registry will enable improved access and utilization of distributed digital content sources.

As part of this development there is a requirement to develop a sustainable identifier infrastructure to persistently identify content across all Jurisdictions. One of the PILIN output is to deliver a value added service called "**Appropriate copy service**". This service determines which instance of a resource is appropriate in what context. Rather than presenting all instances to the end user for browsing, the service determines which instances to the end user for browsing; the service determines which instance the end user should be directed to.

This presents a very strong business case for TLF and PILIN to collaborate and develop this service for mutual benefit. It is proposed that a 3 months pilot be commissioned effective April 2007.

13.3.10.1 About TLF

The Le@rning Federation Initiative is an initiative by all education Ministers in Australia and New Zealand to collaborate in developing online curriculum content and delivering it electronically.

Between 2001 and 2006, it produced extensive online content in a range of priority curriculum areas through partnerships with multimedia developers and cultural organisations. This content, which is hosted in a sophisticated content repository, is highly valued by teachers.

The Le@rning Federation has focused attention on broader policy issues of technical interoperability, intellectual property sharing, digital copyright, bandwidth provision for schools and the importance of implementing Information and Communication Technologies (ICT) across all education sectors to support the growing innovations, enterprise and knowledge economy priorities.

In May 2005, MCEETYA endorsed a proposal to further invest in The Le@rning Federation for 2006–2008 with a specific set of targets. These include a further 4 000 items of online content; framework, standards and structure for the sharing of online curriculum and to enable sharing and peer-reviewing of content developed by teachers; arrangements with vendors to support distribution and use of content; and support for a local education digital content industry.

13.3.10.2 Collaboration background and opportunity

The Le@rning Federation Phase Three Plan introduced the concept of a central registry and metadata repository which would enable Australian and New Zealand school educational jurisdictions to provide enhanced discovery mechanisms for their teachers. This registry will enable improved access and utilization of distributed digital content sources.

- The issues addressed by registry are listed as follows:
- Sharing, re-use and leveraging of content
- A broader range of digital resources from cultural institutions
- Quality assured content
- Interoperability between systems
- Interoperability of Metadata

- Content identification
- Persistence of content and access
- Discovery improvements
- Statutory licence fees
- Bandwidth aggregation opportunities

In order to address some of the above issues a sustainable identifier infrastructure is required by The Le@rning federation.

PILIN project aims to meet the identifier infrastructure needs common across educational, research and cultural communities, with a focus on supporting e-learning and e-search.[2]

Due to the nature of PILIN project and TLF's registry requirements, there exist an opportunity for both the TLF and the PILIN to work together to solve a common problem.

This proposal outlines this collaboration opportunity and provides a basis for further discussions to be held in this joint activity.

13.3.10.3 About TLF requirements

The Le@rning Federation (TLF) achieved its Phase Two goal to collaboratively develop and make available a continuing supply of digital educational content in priority areas agreed by AESOC. As a result, amongst the key infrastructure achievements were:

- a sophisticated content management system (the Exchange) which facilities online content development processes and enables content to be distributed to jurisdictions;
- an intellectual property rights management tool (CRISP) to support the licensing of third party content; and the adoption of specifications that enable the online content to be delivered successfully throughout the distribution chain.
- K-12 education jurisdictions within Australia and New Zealand are now successfully obtaining content from TLF and distributing it to schools. In addition, many jurisdictions have also either developed or procured online content from other sources that are made available to teachers. Tools and systems have also been developed locally to support these processes.

This current environment has resulted in a situation where a piece of content is being referred to by different people, by different names and by different methods. With the development of TLF Registry there arises a need to refer to this content in a manner so that a user can be directed to a piece of content that is located in area where they have access to. Moreover this need to happen in a fashion that user is not aware of complexity in providing the right content from right location but they just get the content very elegantly.

As a consequence, there is a need for the content to be persistence, easily identified and hence a sustainable identifier infrastructure is required.

13.3.10.4 PILIN – Deliverables and Value added services

One of the PILIN output is to deliver a value added service called "**Appropriate copy service**". This service determines which instance of a resource is appropriate in what context. Rather than presenting all instances to the end user for browsing, the service determines which instances to the end user for

2 Persistent Identifiers and linking Infrastructure (PILIN) Project Summary

browsing; the service determines which instance the end user should be directed to. Choices could be based on accessibility, location, version, licences, FRBR category…..[3] This "Appropriate copy service" seems to be addressing the TLF's need as discussed in section 3

13.3.10.5 Scope of collaboration

TLF proposes that a pilot be undertaken in collaboration with PILIN

- To develop this Appropriate copy service and test if this service can address the TLF needs of identifying and referring to content in a standard fashion across all jurisdictions.
- To elegantly redirect users to the location where they are eligible to receive content from.

13.3.10.6 Proposed Timelines

TLF timelines for Registry project are as per the following table:

• Circulation of draft Concept of Operations document to the Le@rning Federation Consultative Committee for comment	**March 2007**
• Discussion of the concept with the Le@rning Federation Consultative Committee and/or others within each ANZ jurisdiction	**March 2007**
• LFCC sign-off on the concept	**April 2007**
• Requirements gathering and documenting	**April 2007**
• Development / system integration of registry pilot	**June 2007**
• Feedback and further requirements gathering and documenting	**December 2007**
• Further development / system integration of full solution	**June 2008**

Based on these timelines following timelines are being proposed for PILIN "**Appropriate copy service**" pilot

• Proposal for the Pilot to PILIN	**April 2007**
• LFCC presentation on pilot concept	**April 2007**
• Pilot budget , project plan agreement and approval	**Early May 2007**
• Project team formation , Role definitions and scope agreement	**Early May 2007**
• Requirements gathering and documenting	**April/ mid May 2007**
• System Development and Testing	**July 2007**
• Proof of concept approval	**Early August 2007**

13.3.10.7 TLF's commitment to the pilot

TLF proposes the following commitment towards the pilot based on funding from PILIN:

Commitment to agreed timelines

- a Business Analyst during the requirements phase
- a Java Developer for PILOT duration
- Support in procurement necessary hardware and software

3 PILIN priorities document

- Project Manager
- Testing and QA resources

13.3.10.8 Next steps:

- The next step from here is for PILIN team to review and accept the proposal formally and advise TLF.
- Once the proposal is formally agreed that project costs and funding arrangements need to be identified.
- Following this, a project team needs to be established and pilot can commence as per timelines and activities defined in

13.4 University of Queensland: Integrating Identifiers into Annotation Harvesting

pilin persistent identifier linking infrastructure

web: http://resolver.net.au/hdl/102.100.272/0N8J991QH

email: policy@pilin.net.au

To cite the latest version of this work use http://resolver.net.au/hdl/102.100.272/K89HHGNQH
To cite this version of this work, use http://resolver.net.au/hdl/102.100.272/K89HHGNQH

Stakeholder	Name of Project	Resources used in terms of time estimate of $	$ allocated out to stakeholder
UQ/DART	Integrating PILIN Identifiers with DART Annotation Harvesting, Aggregation and Searching	Fund up to a total value of $50K including PILIN team time and external costs	$45K

13.4.1 Description of Project and Objectives

As part f the DART project, the University of Queensland has developed a harvesting interface to a server for Annotea annotations, allowing community annotations of research and other data to be harvested by a repository, and incorporated into the searchable metadata for the resource being annotated. The community annotations are to remain distinguished in discovery from the objective metadata provided by the hosting institutions.

In light of this project, the University of Queensland have engaged with the PILIN project to investigate how persistent IDs contribute to the annotation harvesting workflows. This includes managing object-to-annotation and annotation-to-annotation relationships. It also includes the display of harvested annotations to users in search results for an e-research collection. Persistent IDs identify the annotations an (where practical) the annotation targets (resources, and other annotations). They are meant to ensure that the annotations themselves, and the relations of annotations to their targets, persist despite changes in the annotation servers, and (where practical) in the resource repositories.

The project also involves the National Library of Australia as a provider of testbed data from Pictures Australia, and of NLA's functional requirements from the process.

13.4.2 Project Responsibility

13.4.2.1 Role of PILIN Team

Nigel Ward: Investigate how persistent ids can help the annotation harvesting process, including managing annotation-on-resource and annotation-on-annotation relationships

Simon Porter: use existing documentation provided by UQ on annotation models, in conjunction with eFramework methodologies, to document an annotations federation model as a Service Usage Model, and produce Service Genres and Service Expressions (to be developed by Simon in consultation with Judith Pearce, NLA and Jane Hunter, UQ).

Simon Porter: Document UQ Secure Server as an eFramework Service Usage Model

Nick Nicholas: Provide advice on persistent identifier usage, review Service Usage Model, propose persistent identifier workflows.

Manish Saroha, Simon Yu: Provide advice and assistance on deployment at UQ of Handle Server and PILIN Java Handle API.

Dennis Macnamara, overall project to project liaison and administration.

13.4.2.2 Role of UQ Team

Jane Hunter: Project Lead

Ron Chernich: Implemented original OAI-PMH interface for annotation server

Imran Khan: Establish data test bed, Search interface, Harvesting interface, Testing harvesting, PILIN Integration

Anna Gerber: PILIN Integration, Annotation setup

13.4.3 Description of Project Activities and Time Frame

Acquire a small collection of images plus metadata from the NLA (Judith Pearce to assist)	Sept 2007
Assign PILIN identifiers to images and ingest metadata into database (Lucene)	Sept 2007
Design appropriate metadata schema (refine existing schema) for annotations	Sept 2007
Set up new annotation server	Oct 2007
Add PILIN identifier generator to the Annotea code for creating annotations	Oct 2007
Generate set of annotations (with PILIN ids) for images and store on annotation server	Oct 2007
Extend OAI-PMH interface to support harvesting annotations with the new schema from the annotation server	Nov 2007
Test periodic harvesting of new annotations (nightly?)	Nov 2007
Extend web search interface over images - via NLA metadata only, annotations only or both	Dec 2007
Build search results presentation interface that distinguishes and ranks results depending on whether they match NLA metadata or community annotations, and the number of matches	Dec 2007

13.4.4 Description of Project Outcomes

- A harvestable and searchable testbed of annotations on Picture Australia data has been

deployed and demonstrated.

- The PILIN project has done knowledge transfer on setting up a Handle server and deploying the PILIN Handle API to UQ.
- The PILIN project has authored a SUM describing the annotation system and its functional requirements. *This will be* available as a post PILIN 2007 product in February 2008
- The PILIN project has proposed workflows for the best use of persistent identifiers in annotation workflows. This includes restricting persistent identifiers in the first instance to harvested annotations themselves and to annotation-of-annotation relations; and treating the persistent identifier as a URL, to minimise disruption to existing systems.
- Based on input from the PILIN project, UQ has integrated persistent identifiers in their testbed.

The NLA provided open access to Image collections harvested from community libraries via Picture Australia. The Co-Annotea system developed within DART enables communities to annotate web resources with descriptive tags, comments, queries or assessments. The NLA wanted to enrich their existing metadata with community annotations and make this commentary available to the broader community, via their existing Web search interface. This involved:

- Determining the optimum method for integrating harvested annotations with the metadata store for the original collection whilst clearly distinguishing*subjective* community annotations from the more *objective i*nstitutional metadata;
- Implementing a search interface over the integrated institutional metadata and/or annotations;
- Building a user interface for ranking, displaying and browsing retrieved results, based on search term matches to both metadata and annotations clearly identifying the source of the metadata/annotation.
- Building a testbed for demonstration purposes based on: a small collection of images and associated metadata extracted from the NLA; an Annotea annotation server with annotations generated using Co-Annotea and an OAI-PMH interface that allowed harvesting of annotations
- a search and retrieval interface that supported searching across an integrated metadata store that holds both the NLA metadata and the harvested annotations; a presentation interface that displays and ranks retrieved results based on matches to both IM and harvested annotations.

It should be noted that the other example in the proposal, (ie the Protein Databank) was not completed as part of this extension project but will be undertaken as a follow up activity.

The extension project also articulated its outcomes to the community as follows:

- Presentation on this work at Clever Collections Nov 28-29 by Jane Hunter
- Submission of short paper to Open Repositories 2008 waiting to hear whether accepted or not

13.4.5 Evaluative Comments

13.4.5.1 Did the project meet its objectives? describe

The project met its objects of a deployed and demonstrated testbed for discovery through community annotations. It has ensured that harvested annotations are identified on the harvesting server through persistent identifiers. It has ensured that annotation targets of harvested annotations are also identified through persistent identifiers. While it did not undertake both the examples outlined in the proposal it did successfully and more fully implement the NLA image data.

13.4.5.2 Did the Stakeholder finish their activity? describe

Yes, Handle identifiers running on a Handle server in UQ were integrated successfully into annotation workflows, identifying images as targets of annotations, and annotations themselves as digital objects to be harvested. This realised at least some of the proposed workflows, and was sufficient as a proof of concept. (See also Unfinished Business.)

13.4.5.3 What lessons were learnt for PILIN and future identifier services?

- In order to introduce persistent identifiers into existing URL-based workflows with minimal disruption, URL encodings of the persistent identifiers are needed.
- Persistent identifiers can be introduced incrementally into workflows to minimise disruption: server-side before client-side (with transformations happening on the server); harvested objects before identifiers embedded within harvested objects; annotation targets (under one s control in PILIN parlance) before resource targets (outside one s control); harvested transformation of metadata (DC) before metadata in native format on server.
- Embedding persistent identifiers in metadata can be used as a type of reverse lookup, to determine whether an object has already been assigned a persistent identifier. (This idea was also explored with the ARCHER project.)

13.4.5.4 What unexpected things occurred?

The engagement between PILIN and UQ was initially scheduled to begin in September. Because engagement was conditional on UQ completing work on the annotation process, the engagement began at the beginning of November.

13.4.5.5 Is there any unfinished business to note?

This work may inform further work of DART/ARCHER. For this project identifiers were assigned to the primary data objects (images) by hand through Web Handle Management Tool. Future work would include integrating JAHDL into data object ingestion workflow, to guarantee automatic assignment of identifier.

In addition, annotations of annotations currently use local identifiers rather than persistent identifiers for their targets (as proposed by PILIN). Future work would include having annotations of annotations use Handles for their targets instead.

More work will be carried out in 2008 using protein data bank but this will be outside of PILIN activities albeit using PILIN infrastructure now embedded at UQ.

The SUM of the annotation system and its functional requirements will be completed in 2008

as a post PILIN related activity.

13.4.6 Summary

UQ have demonstrated the harnessing of community contributions to metadata in the discovery capability of host repositories; this promises to make repositories more responsive to the needs and expectations of their user communities. Together with PILIN engagement, UQ has made significant progress in making community metadata as persistent managed as the source institutional metadata: this makes the harvesting repository trustworthy as a long-term store of community input on resources, ensuring that such community input is maintained as appropriate. This has been a significant achievement.

13.4.7 Appendices

1. Proposal: Integrating PILIN identifiers with DART Annotation Harvesting, Aggregation and Searching
2. Preliminary thoughts on how persistent identifiers can be applied to the creation harvesting of Annotations
3. DART/UQ Implementation Report

13.4.8 Appendix 1 - Project Proposal

Title: Integrating PILIN identifiers with DART Annotation Harvesting, Aggregation and Searching

Background:

Towards the end of the DART project, work was undertaken at the University of Qld, developing an OAI-PMH interface to an Annotea annotation server. This software module enables community annotations (or most recent updates to the annotation server) to be periodically harvested as XML data via HTTP. This functionality provides an efficient, scalable mechanism by which community annotations of digital resources or scholarly documents can be retrieved by the organization or institution hosting the repository of resources or documents. The annotation server captures community knowledge about a specific collection and the OAI-PMH interface retrieves this information in a standardized format.

However, the work done to date is of little value, without implementing a number of complementary components:

- Using PILIN identifiers instead of URIs for the resources as well as the annotations
- Exploiting the harvested community metadata by making it accessible to the repository search engine and hence to users searching the collection. In order to do this, the harvested annotations need to be mapped to the collection s metadata schema, ingested in the repository metadata store, whilst still being distinguished as annotations and clearly attributed to an individual. The search engine and search and retrieval interface must be modified to support: searching across institutional metadata (IM) only; annotations only; or both. Results of matches with IM must be clearly distinguished from matches with annotations and the retrieved results ranked accordingly.

Research and implementation issues that need to be resolved include: designing an extensible metadata schema to support both annotations and IM; tagging and filtering problematic annotations; managing threaded annotations and deleted annotations; ranking and displaying retrieved results; incorporation of FOAF social networks for ranking matches to annotations.

Example Scenario 1: The Protein Databank provides open access to an archive of published protein crystallography structures. The AnnoCryst system developed within DART enables communities to annotate these structures with descriptive tags, comments, queries or assessments. The Protein Databank administrators would like to enrich their existing metadata and make this commentary available to the broader crystallography community, via their existing Web search interface.

Example Scenario 2: The NLA is digitizing Patrick White's archives to make them available to literary scholars. They would like to also provide annotation services to specific scholarly communities so they can attach and share annotations and specify those annotations that can be harvested and shared with the public, through the NLA web site/search engine.

Objectives:

- Use persistent identifiers to uniquely identify both the resources and the annotations;
- Determine the optimum method for integrating harvested annotations with the metadata store for the original collection whilst clearly distinguishing subjective community annotations from the more objective institutional metadata;
- Implementing a search interface over the integrated institutional metadata and/or annotations;
- Building a user interface for ranking, displaying and browsing retrieved results, based on search term matches to both metadata and annotations clearly identifying the source of the metadata/annotation

- Model annotation services and architectures within the eFramework

Team

Nigel Ward - investigate how persistent ids can help the annotation harvesting process, including managing annotation-on-resource and annotation-on-annotation relationships

Simon Porter – use Jane's existing paper on annotation models and eFramework methodologies to document an annotations federation model as a Sum and produce Genres and Expressions (to be developed by Simon in consultation with Judith and Jane)

Judith Pearce – provide collection of images and metadata for testbed. Provide input outlining NLA's requirements.

Jane Hunter and Imran Khan at UQ - build a proof of concept annotation harvesting service over a sample collection (e.g., NLA Picture Australia images? – to be provided by Judith)

Schedule and Deliverables: September 1- December 31, 2007

Acquire a small collection of images plus metadata from the NLA (Judith Pearce to assist)	Sept 2007
Assign PILIN identifiers to images and ingest metadata into database (Lucene)	Sept 2007
Design appropriate metadata schema (refine existing schema) for annotations	Sept 2007
Set up new annotation server	Oct 2007
Add PILIN identifier generator to the Annotea code for creating annotations	Oct 2007
Generate set of annotations (with PILIN ids) for images and store on annotation server	Oct 2007
Extend OAI-PMH interface to support harvesting annotations with the new schema from the annotation server	Nov 2007
Test periodic harvesting of new annotations (nightly?)	Nov 2007
Extend web search interface over images - via NLA metadata only, annotations only or both	Dec 2007
Build search results presentation interface that distinguishes and ranks results – depending on whether they match NLA metadata or community annotations, and the number of matches	Dec 2007

13.4.9 Appendix 2 - Preliminary thoughts on how persistent identifiers can be applied to the creation harvesting of Annotations

Overview

This paper considers how persistent identifiers can be applied to annotations stored in an Annotea annotation server.

It considers a process for embedding a persistent identifier as an extra element in an annotation schema. The persistent identifier can then be used in the harvested record of the annotation.

Additionally, this paper considers the additional work required to create annotations that refer to other annotations via a persistent identifier.

Discussion

The desired outcome is to have persistent identifiers, rather than local URLs, as the scaffolding on which annotation relies. There are various stages through which this change can be implemented, with increasing effect on annotation processes.

- Annotations are identified for harvesting by PI.
- Annotation targets are identified in harvested record by PI.
- Annotation targets are identified on server by PI.
- Annotation targets are identified on clients by PI.

We think changing annotation clients to rely on PI infrastructure is unrealistic: clients are typically not under central control, and are written only to comply to an interface standard (which requires URIs without requiring PIs). So we propose that only the server changes its handling of annotations and annotation targets.

Moreover, functionality on the server relies on existing local URLs. This is not persistent in the longterm: if the server moves, none of the annotations of annotations links will work any more. But changing the link infrastructure immediately will prove disruptive. For a least disruptive solution, we propose:

- Add PIs as an extra field in the annotation schema, rather than overwrite the local URL.
- NOTE: this can be done asynchronously, in batch mode: it does not need to happen when annotations of annotations are deposited.
- When transforming the annotation records for harvesting, replace the URI in the transformed record with the PI.
- Treat the PI as a URI, to prevent conflict with existing interfaces. (In this instance, the URI for a Handle is a resolution service call.)
- For all annotation targets in the annotation server (i.e. annotations of annotations), change the target local URL to a PI URL. This is the only way to ensure persistence of annotations of annotations on the server (where the targets are still under our control).
- NOTE: this can be done asynchronously, in batch mode: it does not need to happen when annotations of annotations are deposited.

The following scenarios are proposed for an annotation service using persistent identifiers.

1. Annotate crystallographic model

- Client submits annotation of model. Annotation target (the model) is identified by URL.
- Server ingests annotation.
- Server creates Handle. PILIN API
- Server points Handle to Local URL of Annotation PILIN API
- Server embeds Handle identifier, as URL (hdl.handle.net/HANDLE), into annotation as distinct field

Steps after ingestion may occur in batch mode.

2. Harvest annotation

- Server transforms annotation record into format suitable for harvesting
  - As part of this: replace URI of item to be harvested with Handle URL from the annotation schema
- Transformed annotation record is harvested through OAI-PMH
- Harvested record now points to PI of annotation, rather than non-persistent URL

3. Annotate annotation (optional)

- Client submits annotation of annotation. Annotation target (an annotation) is identified by URL. Annotation target resides on same server as annotation.
- Server ingests annotation.
- Server creates Handle. **PILIN API**
- Server points Handle to Local URL of Annotation **PILIN API**
- Server embeds Handle identifier, as URL (hdl.handle.net/HANDLE), into annotation as distinct field

Steps after ingestion may occur in batch mode.

- Server determines that target of annotation resides on the same server, and has a Handle
  - The server can do this by resolving the URL locally, and looking up the schema
- Server overwrites annotation target URL with annotation target Handle URL

This part of the scenario is optional (but needed for true persistence), and presupposes that all annotations on the server already have Handles. This process can also be done in batch mode (Scenario 4)

4. Globally update annotations of annotations to persistent identifiers (retrofitting) (optional)

- Assign Handle to all annotations on server PILIN API
- Embed Handle in schema of all annotations on server
  - For each annotation on server:
  - Is its target an annotation on the same server?
  - Then the target has a Handle, which can be looked up in the annotation schema
  - Overwrite the target URL with the target Handle URL

13.4.10 Appendix 3 – Implementation Report (11 December 2007)

Authors: Imran Khan (initial version), Anna Gerber (update)

This report outlines how each of the deliverables specified in the original PILIN DART Annotation Harvesting, Aggregation ad Searching specification document. It will outline at high level how each of the deliverables has been implemented and where relavant resources can be found.

1. Acquire a small collection of images plus metadata from the NLA (Imran Khan)

For this task, we retrieved a set of images from the picture Australia website related to Architecture. We then used a simple approach of scraping the metadata contained in the HTML search results to gain access to the images metadata. A simple utility application was written for populating the relational database with the images as well as the image metadata.

We also got into contact with the NLA through Judith Pierce about acquiring a set of images in an interesting field. The NLA made available a set of images related to the "Federal Parliamentary Party ot the Northern Territory, April - May, 1912". This collection consists of 133 photogrpahs and there is an OAI interface for the NLA available at http://www.nla.gov.au/digicoll/oai. It maybe usefull in the future to use this dataset instead of the existing one.

2. Assign PILIN identifiers to images and ingest metadata into database (Lucene) (Imran Khan, Anna Gerber)

At present images and their metadata are stored in MySQL database and PILIN identifiers were assigned to the images manually through the PILIN WHMT and the identifiers were added to the database. There is no use of Lucene in this initial version. This is not an issue with such a small set of data but it maybe necessary in the future to revert to Lucene as the way for indexing the DC metadata about the images. For a larger set of data, JAHDL should be used to allocate PILIN identifiers during the ingestion rather than using the web management tool.

3. Design appropriate metadata schema (refine existing schema) for annotations (Imran Khan, Anna Gerber)

This task involved modifying the exisiting Annotea schema to handle keywords. Essentially all that has been done is that a new OWL Object property with domain Annotation and an open range has been added. The range can be any instance of a class, identified by a blank node.

In doing this it is possible to assign any Ontological term to an annotation. For the purposes of the PILIN/DART project, we are using the terms to those defined in the AustralianArchitecture custom ontology. This ontology is the default ontology used by the IE Annotea sidebar for the creation of Semantic annotations, however the ontology used by the Sidebar can be changed in the preferences. The AustralianArchitecture ontology is defined in OWL DL and has been defined by Brendan Mauger (UQ). This ontology is publicly available from http://metadata.net/pilin/AustralianArchitecture.owl.

One other minor refinement of the Annotea schema is definition of an additional subclass of the Annotation type called Feedback. This subclass is essentially the same as a Comment, but is a refinement of the Comment which is used to provide feedback on some content in the

form of free text. An identifier element was also added to the schema, to represent the persistent identifier of the annotation.

4. Set up new annotation server
(Anna Gerber)

The Annotation server is running at http://maenad.itee.uq.edu.au:8080/Annotea

The pictures repository is running at http://maenad.itee.uq.edu.au:
8080/pilin/au.edu.uq.itee.eresearch.pilin.gwt.Main/Main.html

A handle server is also running on maenad.itee.uq.edu.au and the web inteface is avilable at port 8000. The handle namespace is 102.100.738

5. Add PILIN identifier generator to the Annotea code for creating annotations
(Anna Gerber)

PILIN identifiers are added to the Annotations in batches (we are using the Quartz API to schedule the identifier allocation). The frequency of allocation is configurable, and is currently set at every two hours. The PILIN idenfiers are created using JAHDL, and are only used in the OAI harvested metadata for external clients to identify Annotations and in the query interface - they do not replace the UUIDs used internally by the Annotation Server. The PILIN identifiers are not currently used by Annotations to identify Annotations that they annotate, however this could be implemented in future. For ease of implementation, the handle suffix for the identifiers is the UUID of the Annotation.

6. Generate a set of annotations (with PILIN ids) for images and store on annotation server
(Anna Gerber)

Some annotations have been created for testing purposes on the maenad server, however, as the server and pictures repository are currently undergoing testing, bugfixes and some minor modifications, the annotations are frequently changed or deleted. It is anticipated that the final set of demonstration annotations will be created in late December.

7. Extend OAI-PMH interface to support harvesting annotations with the new schema from the annotation server
(Imran Khan)

This task involved the modification of the existing annotation server to handle keywords included in annotations. This involved modifying the OAI cat interface so that it will retrieve the term elements from the Annotation RDF and return the type of the instance which is associated with this property back as an DC subject element.

8. Test periodic harvesting of new annotations (nightly?)
(Imran Khan)

The PILIN Picture website makes use of the quartz API to be able to schedule harvests. This configuration can be found under src/java/conf/pilin.properties. It is possible to configure by the intervals in minutes and seconds of how often to perform a harvest. Currently the service harvests every 10 minutes, however it it is probably better to set it to a higher value.

The default spring configuration only makes use of a single OAI interface for harvesting. However it is possible to add additional repositories by modifying the spring configuration file and specifying additional repository locations.

9. Extend web search interface over images - via NLA metadata only, annotations only or both
(Imran Khan, Anna Gerber)

This feature is supported. It is possible to searches over all of the metadata or a subset of DC elements (e.g. subject, or ontology keywords from the semantic annotations only). The underlying database structure is currently being reviewed and may be replaced with Lucene indexes.

10. Build search results presentation interface that distinguieshes and ranks results depending on whether they match NLA metadata or community anntoaitons, and the number of matches
(Imran Khan)

This feature is not fully tested. The current user interface does not use any pagination but does display the number of results returned for a given query. In terms of ranking results, Imran has tried to use order results by whether or not they are local or not through the HQL (Hibernate Query Language). However as yet this feature has not been fully tested, probably requires more work.

Other developers involved:

Ron Chernich (implemented the original OAI-PMH interface in the Annotation Server)

13.5 LORN (in conjunction with education au) VET Sector
http://lorn.flexiblelearning.net.au/Home.aspx

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
V0.1	Yyyy-mm-dd	DRAFT: Initial draft for stakeholder discussion.	PILIN/???? hdl:102.100.272/TBA
V0.2	Yyyy-mm-dd	Changed fonts	PILIN/TBA hdl:102.100.272/TBA
V1.0	Yyyy-mm-dd	RELEASE: Initial release to public	PILIN/TBA hdl:102.100.272/TBA

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

Stakeholder	Name of Project	Resources used in terms of time estimate of $	$ allocated out to stakeholder
LORN with Education au **VET sector**	Digital Objects, derivatives and multiple concurrent versions	Fund as an Extension Project with a commitment of up to $15K of PILIN team time. Key VET stakeholder, relationship service useable elsewhere including e research	Nil

13.5.1 Description of Project and Objectives

The Australian Flexible Learning Framework has undertaken to pilot a persistent identfier system for the LORN repository of learning objects. This pilot follows from a study already made by Flexible Learning of the business case for persistent identifiers in LORN. Beyond the usual requirements of creating, maintaining and resolving persistent identifiers, LORN also requires an identifier infrastructure for value-added services it needs to provide its community. LORN repositories store multiple adaptations and customisations of the same original content to different disciplines, locales, environments, and jurisdictions. A relationship service is crucial to discovering content related in this way, and this service is dependent on identifiers not specific to a specific instance of a file. Persistent identifiers are also necesary for the successful association of digital rights enforcement and digitial rights management with content.

The engagement project aimed:

- to deploy persistent identifier infrastructure that can be integrated with LORN management and discovery tools;
- to develop a relationship service based on persistent identifiers, which can be used to discover related content on LORN;
- to explore the use of persistent identifiers for digital rights management on LORN.

13.5.2 Project Responsibility

13.5.2.1 PILIN

PILIN aimed to contribute persistent identifier infrastructure to Flexible Learning which Flexible Learning could arrange to be integrated into LORN. It is also helped to guide the devlopment of a relationship service through a prototype developed in PILIN, and made recommendations of service structure and architecture.

13.5.2.2 Stakeholder

Stakeholder Contacts were Owen O'Neill at LORN, Jerry Leeson at Education au. The LORN team aimed to use and test the PILIN functionality inside the Learning Object Repository Network for tracking digital resources and derivatives/versions of resources.

13.5.3 Description of Project Activities and Time Frame

PILIN contributed:

- Expertise on deploying identifier management systems; (May)
- APIs for integrating persistent identifier management, resolution and querying with LORN tools; (July)
- Tools for managing persistent identifiers (Web Handle Management Tool); (August)
- A FRBR-based relationship service, as a prototype of the relationship service to be developed by LORN (October)
- Guidelines for developing identifier policy; (September-October)
- Suggestions on modelling relations between objects, to be represented in a relationship service. (March, October)

A series of meetings took place throughout 2007. PILIN team members met with both LORN technical staff and the wider LORN National reference group. PILIN provided software toolkits for use by LORN. The LORN contact, Owen O'Neill also participated in related e framework developments with PILIN team members.

Towards the end of 2007 LORN technical staff road tested some of the documentation of the PILIN software outputs.

13.5.4 Description of Project Outcomes

LORN has adopted Handle as their preferred approach to persistent identifier implementation. They will use the PILIN software outputs and the policy guideline documents in their 2008 full implementation. LORN is a pivotal project within the VET sector in Australia. The adoption by LORN of Handle and their keen recognition of the essential nature of persistent identifiers is therefore a significant penetration of the VET sector. LORN (and E Standards, a sister activity under the Flexible Learning initiative, http://www.flexiblelearning.net.au/flx/go would be ideal "outreach agents" for the VET sector for the proposed national persistent identifier service.

13.5.5 Evaluative Comments

- **Did the project meet its objectives**

Engagement of the LORN community with PILIN, and demonstrations of both PILIN outputs and the pilot study already undertaken by Flexible Learning, have convinced LORN of the value of a persistent identifier infrastructure, and LORN is committed to going ahead with persistent identifiers as an infrastructure.

- **Did the Stakeholder finish their activity**

The policy and information modelling required by LORN's use of persistent identifiers has been advanced through engagement with PILIN. PILIN's thinking about persistent identifiers has been challenged and enhanced through engagement with the LORN project. In particular LORN's approach to derivative works and multiple concurrent versions has informed the FRBR software development. It has also been useful in discussions with the e- research community which has similar requirements. The full implementation of persistent identifiers within LORN will now occur in 2008.

- **What lessons were learnt for PILIN and future identifier services**

The extension project focused PILIN attention on the range of relationships that could be modelled between identifiers early in the project. Although PILIN has implemented and used the FRBR model as only a very generic relationship service, this work can still be used for more specific services. The exercise has highlighted the importance of deciding on an information model before they start using persistent identifiers for their objects, which has become one of PILIN's core recommendations to projects.

The relative timing of creating an object with an embedded identifier, and assigning an identifier to the object, has emerged as an issue in LORN use of identifiers: identifiers need to be created before they are embedded, but usually their association cannot be registered until after the object is fully populated and registered itself. This has highlighted the need for incompletely populated ("temporary, reserved") identifier objects to be registered, which refines the workflow for identifier registration and the model of identifiers behind it.

- **What unexpected things occurred**

Delays to deploying a developer from Education au to work with LORN to implement persistent identifiers have pushed full implementation back to 2008. Nevertheless valuable engagement occurred for all three parties.

- **Is there any unfinished business to note**

The implementation stage of the project, and the detailed information modeling of possible

relations between learning objects, will occur in 2008. LORN are very keen to work with a National Persistent Identifier Utility service in 2008.

13.5.6 Summary

The LORN engagement project demonstrated a strong commitment to persistent identifier infrastructure in the user community. It also highlights the importance of identifier providers supporting value-added services, which may lie outside the strict domain of the identifier management system.

13.5.7 Appendix

Appendix 1: Persistent identifier business case and pilot study recommendations. www.flexiblelearning.net.au

13.5.8 Appendix 1: Persistent identifier business case and pilot study recommendations. www.flexiblelearning.net.au

13.5.8.1 Introduction

A persistent identifier is a **unique** and **permanent** identifier which is assigned to a digital object. It is possible that an object may be assigned more than one persistent identifier but an identifier can only refer to one object. The term "persistent identifier" as used in this document refers to the concept, not an actual system, policy or implementation.

A persistent identifier never changes and is not affected by its object being moved or renamed: using an object's persistent identifier instead of its URL means that the link will never break. This persistence can be used to facilitate DRE/DRM (Digital Rights Expression/Management), e-commerce transactions and facilitate better management of resources.

13.5.8.2 Problem statement

There is currently no universal system for allocating and managing globally unique persistent identifiers for digital objects (usually learning objects, but could be any digital object) in the VTE sector. This has the potential to cause a number of issues including:

- No definitive way of distinguishing between a localized version of an object and the original.
- Association of digital rights to the object may be ambiguous (to which versions of an object do particular DRM requirements apply?).
- When an object is moved, deleted or archived, links to that object break

Identifiers for objects are currently managed at the database level and there is no coordinated approach between repositories.

13.5.8.3 The case for a persistent identifier system

Persistent identifiers provide an underlying foundation for developing, supporting and enforcing policies around the management and administration of objects in a collection. In a practical sense, the persistent identifier will be used by **systems** to uniquely identify a specific digital object and by **people** when linking to or referring to an object (it works in the same way as a standard URL when used in a browser). Issues identified in this section will become magnified as a collection of objects grows in size. Section 7 of this paper contains a number of use-cases demonstrating the purposes of persistent identifiers.

13.5.8.3.1 Customising digital objects / versioning for digital objects

If objects are customised and added back to the LORN, a unique way of identifying objects is vital to clearly distinguish between the original and customisations. A persistent identifier can be used to unambiguously distinguish between different versions of the same object.

13.5.8.3.2 Changing the location of digital objects

From time to time, digital objects may need to be moved from one location to another. For example, if an organisation is taken over by another or if two repositories are merged, these changes result in the URLs (ie. location) of digital objects changing. Such changes cause broken links that are difficult to fix and makes content less reliable.

13.5.8.3.3 Associating DRE/DRM to objects

When allocating digital rights requirements to an object, it is very important to be able to clearly and unambiguously identify the object with a digital rights statement. To track an object or enforce digital rights requirements, a unique way of identifying it from any other is vital. For example, where there are customised versions of an object, the digital rights expression could be used to clearly identify the original as the only officially supported version.

13.5.8.3.4 Deleting/archiving digital objects

When an object is deleted from a database, all links to that object break. This is a perennial problem for web content. A persistent identifier facilitates the establishment of processes for deleting and/or archiving objects such as providing information to the user about why the object was removed, suggesting an alternative or automatically redirecting to an alternative. As LORN repositories mature, regular archiving of objects will become a more common occurrence.

13.5.8.4 The case for using the "Handle System"

In the persistent identifier paper [1] discussed by the E-standards Expert Group in May 2006, a number of standards for persistent identifiers were considered including:

- The Handle System
- A generic naming approach
- Digital Object Identifier (DOI) system
- Persistent Uniform Resource Locator

The Handle System [2] was selected as the recommended choice for the pilot study after discussion by the E-standards Expert Group and consultation with a number of external organisations. The Handle System provides **services** for **allocating**, **managing** and **resolving** identifiers. The resolution service resolves an object's identifier with the actual location (web address) of that object. Reasons for recommending the Handle System include:

- The Handle System forms the basis of other options including DSpace [3] and the DOI system [4]
- It was selected by the ARROW project (a network of Australian research repositories) [5] after extensive consideration of a number of alternatives
- It is freely available and has been implemented by education.au
- It is the preferred basis for a national system proposed by a number of organisations such as ADL Australia and IMS Australia
- The Handle System can be used with existing legacy systems

13.5.8.4.1 Success measures

The persistent identifier system would be required to support/enable these key processes.

Key Processes/Services
The system can automatically allocate a globally unique persistent identifier when an object is added to the network

The resolution service automatically resolves the URL of an object from a persistent identifier
If an object is uploaded with an existing PI, the user is warned and the adding process is halted (a policy is required to determine when a new PI should be assigned to new/localised versions)
When an object's URL is changed (eg. the location of the object changes) the resolution service is automatically updated to reflect this change.
Updates to URLs in the resolution service should be able to be done individually or as batch processes (in groups)
When an object is deleted, the persistent identifier of that object should be able to point users to some useful information such as possible alternatives and/or a reason why the object was deleted
A persistent identifier can be used to associate DRE to an object
The persistent identifier can be used for all linking to an object
The use of persistent identifiers involves minimal modifications to legacy systems (particularly repositories and databases)

13.5.8.5 Recommendations for pilot project

A pilot project is recommended in order to trial persistent identifiers in a realistic environment. This pilot should be conducted within LORN. The pilot persistent identifier system (software applications and policies) should support or enable all key processes/services outlined in Section 4.1 (Success measures).

13.5.8.5.1 Pilot project outcomes

Objective	Description
Practical experience in persistent identification	The project will deliver experience in allocating, managing and using persistent identifiers for digital objects
Evaluation of the Handle System	An evaluation of the Handle System and its advantages/disadvantages to the VTE sector
Support VTE negotiations in any proposed national system	The experience from this pilot study will give us knowledge of what the requirements for the VTE sector are for persistent identifiers
Stronger links with key organizations	Working with organisations such as EdNA and TLF in this area should deliver cooperative benefits.
Understanding of the policy issues which underpin this area	A practical implementation will give a clear picture of the policy implications of persistent identifiers
Recommendation on how to implement the system across the network	If appropriate, the pilot should indicate how a persistent identifier system might be implemented more widely
Feedback on the suitability of the selected technologies and software	Technologies and software for using and managing persistent identifiers will be trialled and its appropriateness will be considered

13.5.8.5.2 Technology environment

Software/technology	Description
The Handle System	The Handle System is a distributed information system that allocates and resolves identifiers for digital resources. The system includes an open set of protocols, a namespace, and a reference implementation of the protocols
DSpace	An open source repository which implements the Handle System

13.5.8.5.3 Assumptions

Assumptions
The pilot is able to use the EdNA DSpace implementation with existing LORN repositories
The pilot system does not require major modifications to participating repositories
Minimum of one repository in LORN willing to participate
Existing database-level identifiers can be used as the basis for persistent identifiers

13.5.8.5.4 Major Milestones/deliverables

13.5.8.5.4.1 Set up of persistent identifier system

Target Date	Milestones/Deliverables	Responsibility
LORN TBA	Set up of Handle System for VTE objects	LORN TBA
LORN TBA	Modification of systems to enable allocation of globally unique persistent identifiers (if applicable)	LORN TBA
LORN TBA	Establish initial policies and usage guidelines for persistent identifiers in the pilot	LORN TBA
LORN TBA	**Deliverable 1: Establishment of pilot system for allocating/resolving persistent identifiers**	**LORN TBA**

13.5.8.5.4.2 Testing of the persistent identifier system

Target Date	Milestones/Deliverables	Responsibility
LORN TBA	System is tested for the following common actions Persistent identifiers allocated to new learning objects Persistent identifiers allocated to localised learning objects Learning objects successfully moved from one repository to another Learning objects successfully deleted/archived	LORN TBA
LORN TBA	System is tested against each use-case in section 7	LORN TBA
LORN TBA	**Deliverable 2: Report on testing of the persistent identifier service**	**LORN TBA**

13.5.8.5.4.3 Evaluation of system and recommendations for implementation

Target Date	Milestones/Deliverables	Responsibility
LORN TBA	Evaluation/suitability of the Handle System	LORN TBA
LORN TBA	Evaluation of the system and suitability for implementation	LORN TBA
LORN TBA	Recommendations on policy	LORN TBA
LORN TBA	**Deliverable 3: Final report on recommendations on suitability of the persistent identifier system and recommendations for implementation**	**LORN TBA**

13.5.8.6 Pilot project evaluation

13.5.8.6.1 Consideration of risk

Risk	Mitigation actions
Repositories may be unwilling to participate due to time/resource constraints	Minimise extra work for repositories Provide additional support to participating repositories if required

13.5.8.6.2 Analysis of alternatives

No action (status quo)	Key factors showing non-viability
No persistent identifier system	Unable to distinguish clearly between different versions of objects, particularly across repositories Links to objects are at risk of becoming broken if objects are moved No support for DRM

Alternative option(s)	Key factors showing non-viability
1. Generic naming approach The National Library of Australia (NLA) has a set of guidelines for generating and allocating persistent identifiers. This uses the principal of an intelligent identifier rather than a dumb number . For example, an oral history sound recording identifier would take the following form: <collection id>-<collection no.>-<series no.>-<tape no.>-<fragment identifier>-<role code> NLA has developed protocols for creating such identifiers.	Identifiers need to be manually allocated, or software needs to be developed to implement this task. As the identifiers are domain-specific, generic software may not be suitable. Such a naming convention would need to be developed within the context of established conventions (DOI, URN, etc) to facilitate interoperability. In-built semantics may be subject to change over time, however existing persistent identifiers can never be changed. For human users, semantics built into an identifier are only useful if the semantics are understood
2. Digital Object Identifier (DOI) System This is an implementation of URI (Uniform Resource Identifier) using the Handle System. The DOI organisation manages the administration of the system and policies for its use, but membership is expensive and there is an ongoing cost ($35,000 USD initial membership fee and annual fee of $20,000 USD. First 5 million DOIs cost $0.01 USD each, afterwards dropping to $0.005 USD per DOI allocated).	DOI also mandates a set of metadata which must be used. Significant cost of membership and ongoing expenses would require significant demand and a very large user base to make the investment worthwhile DOI policies would need to be carefully analysed to ensure they could accommodate any specific requirements of the Australian VTE sector

3. PURL (Persistent Uniform Resource Locator) This is considered an interim measure until URNs are more widely accepted. A PURL points to an intermediate resolution service, rather than the actual object, thereby allowing the object to be moved without its PURN changing.	Not designed as or considered to be a long-term solution

13.5.8.7 Case studies

13.5.8.7.1 Customising / versioning digital objects

13.5.8.7.1.1 Customising a learning object

Narrative

A teacher requires a learning object for demonstrating first aid in commercial kitchens. She finds a learning object originally developed to teach first aid in the workplace and the digital rights associated with the learning object state that the object can be localised. Much of the original information is relevant but she modifies the learning object to make it more specific to a commercial kitchen setting. The teacher now uploads the new version back into the repository so that others can use it, modifying the metadata record and digital rights statement[4] to reflect the changes. The repository allocates a new persistent identifier to the object which unambiguously distinguishes it from the original.

Primary actors

- Teacher/instructor

Stakeholders and Interests

- Teachers Creating, using and modifying learning objects
- Repository administrators who maintain repositories of well described and accessible learning objects

Preconditions

- The teacher finds a learning object that can be customised to meet the needs of her students.
- The learning object is allowed to be modified according to the DRE associated with it.

4 For simplicity, the use-case assumes that the copyright will remain with the original owner, but in practice this may vary, and will depend on the digital rights associated with the original object as well as any digital rights policies.

Main success scenario

- A learning object is found in a repository and customised for use in a different context
- The learning object is added back into the repository
- The new object is allocated a new persistent identifier and modified metadata to distinguish it from the original

Alternative path

The user allocates a meaningful persistent identifier to the object manually (rather than a persistent identifier being allocated automatically).

13.5.8.7.1.2 Versioning for learning objects

Narrative

Learning objects about traffic safety are produced as part of an e-learning project. In NSW a change in traffic safety law requires a slightly different version of the learning object which is relevant for NSW students only. The new version of the learning object is produced with its own persistent identifier and metadata to reflect its usage. This is an instance of two versions of a learning object which are both correct within their jurisdictions.

Primary actors

- Instructional designer

Stakeholders and Interests

- Teachers using learning objects with their students
- Repository administrators who maintain repositories of well described and accessible learning objects

Preconditions

- Two versions of the same learning object are required

Main success scenario

- A learning object is found in a repository and customised for use in a different context
- The learning object is added back into the repository
- The new object is allocated a new persistent identifier to distinguish it from the original

Alternative path

An error is found in the learning object and it is fixed by updating the existing object without creating an alternative version (the object's persistent identifier does not change).

13.5.8.7.1.3 Changing the location of digital objects

Narrative

The learning objects in two repositories are being moved to a single repository because two TAFE institutions are merging. As a result, the URLs of each learning object in the repository will change. Because all web links to learning objects in the repositories are made using persistent identifiers, all links to the learning objects will continue to work after they are moved to the new repository.

Primary actors

- Repository administrators

Stakeholders and Interests

- Teachers use learning objects and require continuity of existing materials which have web links to learning objects.
- Students Use learning materials and expect high quality learning materials that don t contain broken links.

Preconditions

- Each learning object has been allocated a persistent identifier

Main success scenario

- Learning objects are always linked using their persistent identifier and not their URL
- The two repositories are merged into one, which reflects the newly created institution
- The resolution service which resolves a persistent identifier to its URL (its actual web address) is updated to reflect the changed URLs.

Alternative path

Objects which do not have a persistent identifier would require HTTP redirections to be set up and maintained to avoid broken links.

13.5.8.7.1.4 Associating DRE/DRM to digital objects

Narrative

A Registered Training Authority agrees to contribute a learning object to a national learning object repository on the condition that it is only used for educational purposes. This requirement is recorded in the rights restrictions using the AESharenet “Free for Education” license which is explicitly linked to the learning object using its unique identifier. The DRM information is associated to the object irrespective of where that object is located. A later cut down version of the learning object is released which is assigned the AESharenet “Unlocked Content” licence. This does not change the digital rights requirement for the first object and the two objects can be clearly distinguished from each other.

Primary actors

- Content contributor/rights owner
- Teachers and students (learning object users)

Stakeholders and Interests

- Repository administrator

Preconditions

- Rights are expressed in the metadata record
- The learning object has been assigned a persistent identifier

Main success scenario

- The digital rights information referred to in the object s metadata is explicitly linked to the object using its persistent identifier
- Users are provided with a clear understanding of the digital rights restrictions associated with the object

Alternative path

The content owner is not confident that his/her digital rights will be respected and decides not to add the object to the repository.

13.5.8.7.1.5 Deleting/archiving digital objects

Narrative

During a scheduled audit of learning objects in a repository, a learning object was identified as needing to be completely removed from the repository and archived as it is out of date and contains information that is now incorrect. During the process of removing the object from the repository, the repository administrator is required to add some information about why the object has been deleted, and provides a persistent identifier to an alternative object. If people are still linking to the out of date object, the link will now resolve to a page with the message advising the user why the object was removed while providing the suggested alternative.

Primary actors

- Archivist, repository maintenance person

Stakeholders and Interests

- Teachers/instructors using objects for teaching and learning
- Repository administrators

Preconditions

- An existing object is deemed to require removal from a repository. The decision to archive may have been made as part of a standard maintenance procedure on the repository

Main success scenario

- An object is identified which requires removal from a repository
- The person removing the object is required to enter a reason for the removal and provide the persistent identifier of one or more alternative objects

Alternative path

If an object is being directly replaced by a new object, a notification to users could be provided with a link to the new object or the persistent identifier could automatically redirect to the new object without notifying the user.

13.5.8.8 References

1. Persistent Identifier Discussion Paper discussed at the E-Standards Expert Group 18 May 2006. Email standards@flexiblelearning.net.au to request a copy.
2. Handle System: Is a system to store identifiers (known as handles) of resources and resolve those handles into the information necessary to locate, or otherwise make use of the resources. http://www.handle.net/
3. DSpace is open source repository software which provides tools for managing digital objects. DSpace also implements the Handle System. http://www.dspace.org/
4. DOI (Digital Object Identifier) is a system for providing and managing permanent identifiers for digital objects. The DOI System is managed by the International DOI Foundation. http://doi.org/
5. ARROW (Australian Research Repositories Online to the World) Project is a DEST funded project. http://arrow.edu.au/

14 Appendix B: Other Stakeholder Engagement Summary

pilin persistent identifier linking infrastructure

web: http://resolver.net.au/hdl/102.100.272/0N8J991QH

email: policy@pilin.net.au

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

14.1 Background

All stakeholders provided valuable input for use cases for the PILIN project. All stakeholders received valuable input from PILIN team members and immediate access to developing PILIN policy documents and software toolkits. They were also able to access help desk support for the software toolkits. All stakeholders were provided with access to the PILIN stakeholder wiki. Most of these stakeholders attended the stakeholder meetings in April and November Role of PILIN Team.

The following stakeholders were involved with PILIN but were not nominated extension projects:

14.2 APSR http://www.apsr.edu.au/

The Australian Partnership for Sustainable Repositories (APSR) project had significant input into the PILIN project and vice versa. APSR provided useful feedback on the Identifier SUMS and vigorous input into the PILIN and FRED stakeholder days. It was planned to run an “out there” extension project with APSR but a change of personnel at APSR derailed this plan

APSR also consulted with PILIN on the application of identifiers to collections and to services, in planning their ORCA Collection Service Registry.

14.3 ARROW http://arrow.edu.au

The Australian Research Repositories Online to the World (ARROW) project is actually the over arching body for PILIN. PILIN is part of the ARROW 2 suite of projects. The PILIN development team, including the business analyst were house at ARROW central at Monash University. PILIN worked closely with ARROW throughout the project and one of the PILIN software toolkits provided a solution for an ARROW use case

14.4 Cooperative Research Centre for Spatial Information http://www.crcsi.com.au/

PILIN met with the CRCSI one and discussed identifier infrastructure. No concrete activities were undertaken, and the CRCSI was unable to attend the stakeholder days.

14.5 Dictionary of Sydney http://www.dictionaryofsydney.org/www/html/7-home-page.asp

The Dictionary of Sydney project has input from Sydney University and State Library NSW. PILIN held several meetings with the Dictionary of Sydney team and provided access to PILIN outputs. They were particularly interested in the policy guidelines, information modelling, and exploring a services view of identifiers.

14.6 Education.au

Education au was mainly a stakeholder via the LORN project and are referenced there. Jerry Leeson from education au attended the two PILIN stakeholder meetings.

14.7 FRED

The Federated Repositories for Education (FRED) project has worked very closely with the PILIN project, sharing personnel (technical manager and business analyst), jointly developing solutions for stakeholders, eg Appropriate Copy for the Learning Federation, running and facilitating stakeholder engagement days.

14.8 Howard Florey Institute (HFI, Melbourne Uni) http://www.florey.edu.au/ and Arcitecta Pty Ltd, http://www.arcitecta.com as the HFI (Vendor)

Discussions were held with the Institute and the vendor, going into some detail on technical considerations and architecture; and access was provided to PILIN outputs. No concrete activities were undertaken.

14.9 I Spheres http://www.ispheres.org/

PILIN met with the I Spheres team twice and provided access to PILIN outputs but no

concrete activities were undertaken.

14.10 National Library of Australia (NLA) http://www.nla.gov.au/

The NLA were represented on the PILIN extended team by Judith Pearce, who was also on the PILIN Steering Committee. Matthew Walker from the NLA also participated on the Stakeholder wiki. The NLA also participated in the University of Queensland Annotation Extension Project. The NLA has a strong track record of developmental work with persistent identifiers and provided valuable input in to the PILIN project. Staff from NLA attended both PILIN stakeholder days. Towards the end of the project NLA was identified as the potential national service point for persistent identifiers. PILIN team members worked with NLA staff to put together a submission to DEST for funding for such a service in 2008. On the basis of the possibility of that service being funded, NLA and PILIN team worked closely to implement knowledge transfer strategies to ensure the NLA were well equipped to initiate such a service in 2008. Activities included installing PILIN software on NLA servers, a Knowledge Transfer Day on 5 December at the NLA with the whole PILIN team present and numerous other visits, phone calls and email exchanges. At one point it was considered that NLA would be a nominated Extension Project but partly because of the 2008 service preparations, it was decided not to go ahead with the extension project.

14.11 Neurosciences Research Group (Medicine, Nursing and Health Sciences, Monash University) http://www.med.monash.edu.au/medicine/mmc/neurosci.html

Discussions were held with the Research Group on identifier scheme alternatives and architectures, and advice was given on the use of identifiers in federated architecture. No concrete activities were undertaken.

14.12 RIDIR http://www.hull.ac.uk/ridir/

The Resourcing Identifier Interoperability for Repositories (RIDIR) project is a parallel project in the UK. The PILIN project manager attended their initial planning meeting in April 2007. PILIN provided input into the RIDIR project planning and also access to all PILIN resources and outputs during the project. RIDIR included all PILIN use cases in its documentation. The RIDIR project does not conclude until April 2008

14.13 RUBRIC http://www.rubric.edu.au/

The Regional Universities Building Research Infrastructure Collaboratively (RUBRIC) project is, like PILIN, managed by University of Southern Queensland. PILIN and RUBRIC have collaborated at many levels including the use of software (ICE), cross support from

team members, sharing of personnel (Peter Sefton) and exchange of documentation. RUBRIC have included PILIN guideline material in the RUBRIC Toolkit.

14.14 TILIS
http://www.talc.com.au/TILIS/tabid/220/Default.aspx

The Totally Intelligent Logistics Inquiry Service (TILIS) is a project to establish shared infrastructure in the form of a web gateway for the transport and logistics industry. It is part of the Industry vertical, The Transport and Logistics Centre (TALC). PILIN management and team members held a number of meetings with TILIS staff at the TALC offices in Sydney. At one stage TILIS were keen to “road test” the PILIN software toolkit in an industry setting. However TILIS were unable to attend the PILIN stakeholder days (nor the FRED one) and were difficult to make any contact with as 2007 progressed.

14.15 Summary

Valuable input was received for the PILIN project from most of these stakeholders. At the same time these projects received useful help and support from the PILIN team.