

Creating provenance supergraphs using pingbacks

The Provenance Sharing Network

Nicholas J Car

MODSIM 2015, 4th December 2015

LAND & WATER
www.csiro.au



Outline

- Supergraphs
- PROV-AQ Overview
 - Pingback messages
 - PROMS' extensions
- Lebo Implementation
- PROMS Server implementation
 - Methodology overview
 - Candidate selection
 - Strategies
 - Status Recorder
 - Testing
- Tools
- An Australian Provenance Sharing Network

Supergraphs

- LANDSAT basic processing at NCI
 - Provenance recorded at NCI
 - LANDSAT product generation by GA
 - Provenance recorded at GA
 - Further derived products made at BoM
 - Provenance recorded at BoM
-
- Total graph:



Supergraphs

- LANDSAT basic processing at NCI
 - Provenance recorded at NCI
- LANDSAT product generation by GA
 - Provenance recorded at GA
- Further derived products made at BoM
 - Provenance recorded at BoM
- Total graph:



- How does the NCI know about GA & BoM's work?

PROV-AQ Overview

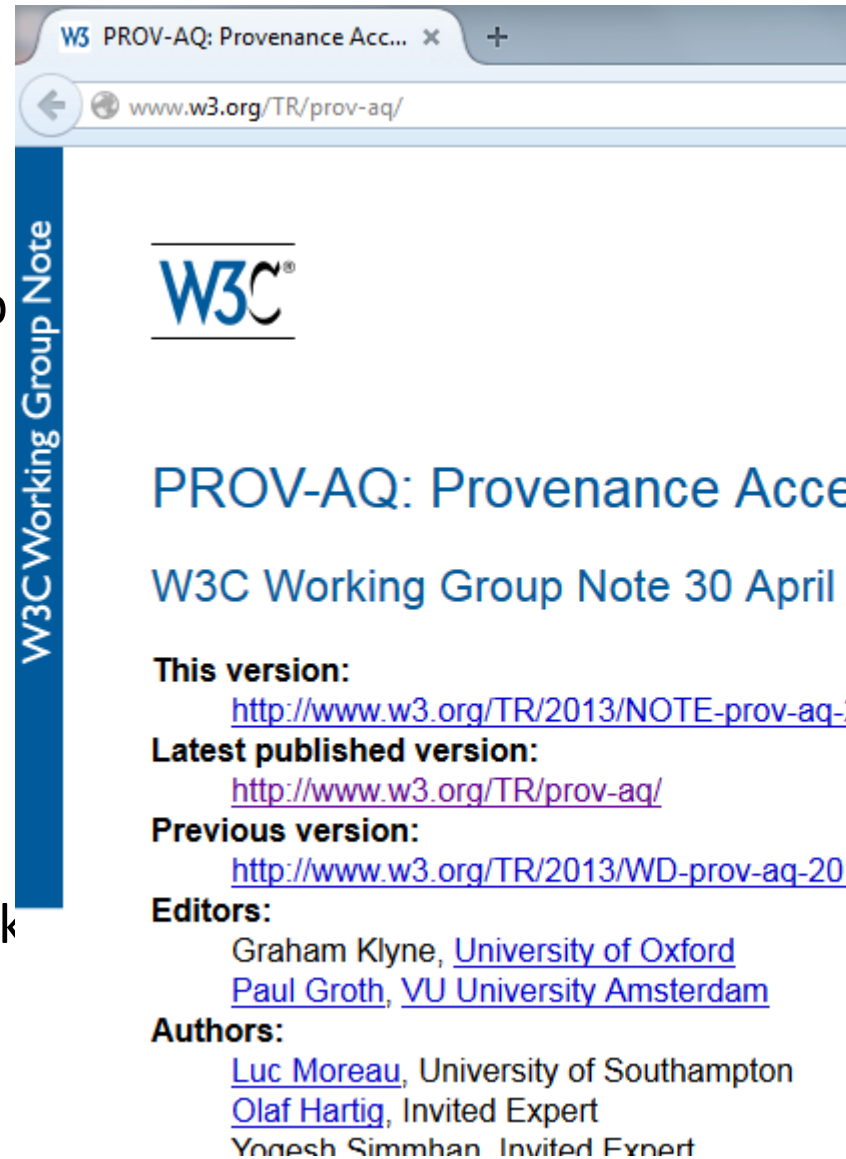
Describes how to locate, retrieve and query provenance information on the web

Uses existing web standards

Part 5. Provenance Pingbacks

- How a resource is used by 3rd parties, communicated back to the publisher

Does not describe how or when a pingback generator or receiver would generate, process or use pingback information.



The image shows a screenshot of a web browser displaying the W3C Working Group Note for PROV-AQ. The browser's address bar shows the URL www.w3.org/TR/prov-aq/. The page features the W3C logo at the top, followed by the title "PROV-AQ: Provenance Access and Query" and the subtitle "W3C Working Group Note 30 April 2013". Below this, the page lists the current version, the latest published version, and the previous version, each with a corresponding URL. It also lists the editors and authors of the document.

W3C Working Group Note

W3C

PROV-AQ: Provenance Access and Query

W3C Working Group Note 30 April 2013

This version:
<http://www.w3.org/TR/2013/NOTE-prov-aq-20130430/>

Latest published version:
<http://www.w3.org/TR/prov-aq/>

Previous version:
<http://www.w3.org/TR/2013/WD-prov-aq-20130410/>

Editors:
Graham Klyne, [University of Oxford](#)
Paul Groth, [VU University Amsterdam](#)

Authors:
[Luc Moreau](#), University of Southampton
[Olaf Hartig](#), Invited Expert
[Yogesh Simmhan](#), Invited Expert

PROV-AQ

Pingback messages 1

- Consist of an
 - HTTP POST
 - Sent to a location related to the resource
 - e.g.: <http://example.com/resource/42#pingback>
 - MIME type text/uri-list indicating locations of more provenance for resource
 - e.g.:

```
POST http://example.org/resource/42#pingback HTTP/1.1
Content-Type: text/uri-list
```

```
http://somewhere.com/some/provenance
```

```
http://some-more.com/other/provenance
```

PROV-AQ

Pingback messages 2

- `prov:has_query_service` message

– e.g.:

```
POST http://example.org/resource/42#pingback HTTP/1.1
Link: <http://somewhere.com/prov-service>;
      rel="http://www.w3.org/ns/prov#has_query_service";
      anchor=http://example.org/resource/42
Content-Type: text/uri-list
Content-Length: 0
```

PROV-AQ

PROMS' extension 1

- Single service endpoint for multiple resources
 - Requires no redirects
 - Use PROV-AQ query service-style format
 - e.g.:

```
POST http://example.org/pingback-service HTTP/1.1
Link: <http://somewhere.com/some/provenance>;
      rel="http://www.w3.org/ns/prov#has_provenance";
      anchor=http://example.org/resource/42
Content-Type: text/plain
Content-Length: 0
```


PROV-AQ

PROMS' extension 2

- Send a whole graph message (PROMS Report with relaxed constraints)

- e.g.:

```
POST http://example.org/resource/42#pingback HTTP/1.1
Content-Type: text/turtle
```

```
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
```

```
<http://placeholder.org#8747df97c3961> a prov:ExternalReport ;
  rdfs:label "ANUGA Busselton Sample"^^xsd:string ;
  prov:nativeId "10"^^xsd:string ;
  prov:reportingSystem <http://localhost:8088> ;
  prov:startingActivity <http://localhost:8088/secure/getJobObject.do?jobId=10>
```

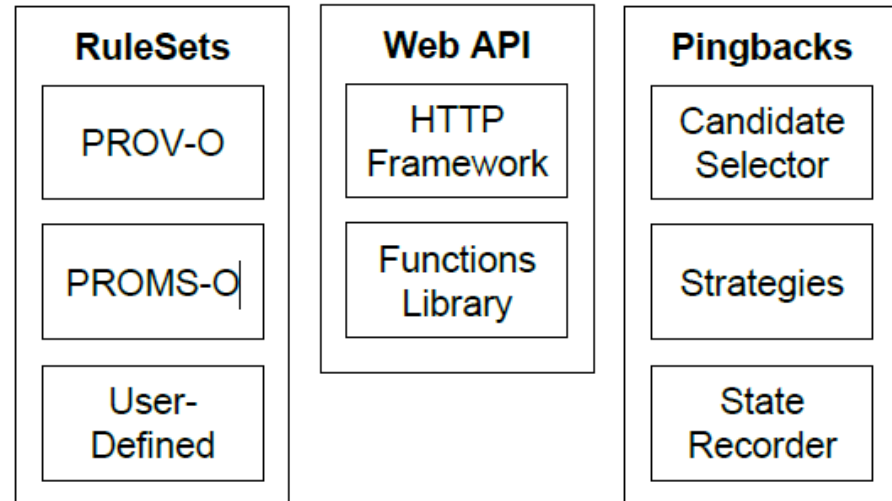
```
.
<http://placeholder.org#e75116a9-ad7d-4701-a85f-3a2204317c3c> a prov:Entity ;
  dcat:downloadURL "http://abc.com/secure/job1234/bustn.asc"^^xsd:anyURI ;
  prov:wasAttributedTo "mailto:everbloom@gmail.com"^^xsd:string
```

Lebo Implementation

- Summary from:
<https://github.com/timrdf/prizms/wiki/Publication:-IPAW-2014>
- Prizms Publishes Host System's prov:has provenance Target
- Prizms Accepts Pingback Pointers
- Prizms Retrieves, Analyzes, and Rehhosts Pingback Pointers
- Criticism:
 - Not generic
 - Not testable
 - Academic only!

PROMS Server implementation

- Automatic if required – just works and shares
- Manual if required – back up
- Can work with non-PROMS systems
- Can work with smart repositories
- Can work according to extensible strategies for different users



A block diagram of PROMS Server's application code.

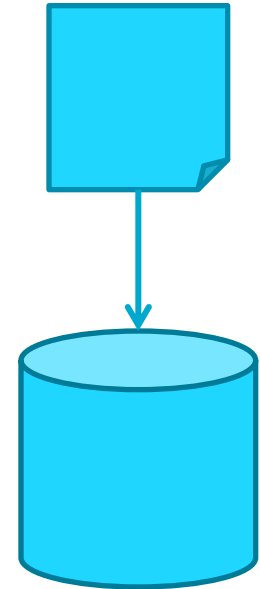
PROMS Server implementation

Methodology overview

When a Report is received by a PROMS Server instance the pingbacks methodology is triggered.

The major steps in the methodology are:

1. determine pingbacks candidate objects;
2. perform pingbacks according to a selected strategy (or more than one);
3. record the success or failure of pingbacks and reschedule pingbacks for failures, according to a retry strategy.



PROMS Server implementation

Candidate selection

No.	Criteria	Reasoning
1	Entity may be a <code>prov:Entity</code> or any subclass of it	Subclasses of <code>prov:Entity</code> , such as <code>prov:Plan</code> or <code>proms:ServiceEntity</code> may be candidates for pingbacks as they can be used as Entities.
2	Entity must be <i>used</i> , not <i>generated</i>	Entities described as having been <code>prov:generated</code> within a Report will have their entire provenance known until another process uses them for another purpose. Therefore, only those not generated by <code>prov:Activities</code> , those linked to <code>prov:activities</code> with a <code>prov:used</code> relationship only will be considered.
3	Entity must not be derived from another Entity	Entities with a <code>prov:wasDerivedFrom</code> relationship to another Entity in the report have their provenance information completely known until used by another process.
4	Entity must not be entirely defined within the Report	Entities defined in a Report with a local URI (usually a hash URI extension to the Report) cannot have pingbacks performed against them as there is no possibility of them existing outside a single Report.

PROMS Server implementation

Candidate selection

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX proms: <http://promsns.org/def/proms#>
SELECT ?e
WHERE {
    # Criteria 1 - inferencing
    #?c rdfs:subClassOf* prov:Entity .
    #?e rdf:type ?c .

    # Criteria 2
    MINUS {?a prov:generated ?e .}
    MINUS {?e prov:wasGeneratedBy ?a .}

    # Criteria 3
    MINUS {?e prov:wasDerivedFrom ?e2 .}

    # Criteria 4
    {SELECT ?b WHERE {
        {?r a proms:ExternalReport .}
        UNION
        {?r a proms:InternalReport .}
        BIND(STRBEFORE(STR(?r), "#" ) AS ?b)}
    }
    FILTER regex(STR(?e), CONCAT("^((?!", ?b, ").)*$", "i"))
}
```

SPARQL 1.1 Query run against
an incoming *Report*

PROMS Server implementation Strategies

No.	Name	Description
0	No Action	Take no pingbacks actions. Implemented by default.
1	Pingback Given	Pingbacks are sent to each Entity with a <code>prov:pingback</code> property defined in the Report.
2	Provenance Given	Bundles of provenance information, not pingbacks, are sent to Entities with properties other than <code>prov:pingback</code> .
3	Known Provenance Stores	PROMS Server is given a list of known provenance stores and attempts pingbacks for each Entity selected by looking them up in those stores.
4	Pingback Lookup	PROMS Server attempts to find pingbacks endpoints for each Entity via Linked Data lookups using each Entity's URI.
5	Provenance Lookup	PROMS Server attempts to discover other provenance endpoints for selected Entities via Linked Data lookups to send provenance bundles to.

PROMS Server implementation Strategies

- Strategy 4 “Pingback Lookup” example:
 1. Dereference the Entity’s URI
 2. Get RDF
 3. Look for declared `prov:pingback` locations
 4. Look for `prov:has_query_service` locations
 5. Look for above locations declared for other metadata objects associated with the Entity
 6. Look for a `dpns:ProvenancePingbackService` that hosts the Entity
 7. Look for... user defined

PROMS Server implementation

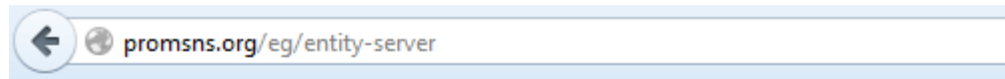
Status recorder

- A DB of Entities for which pingbacks have been attempted, so far unsuccessfully
- Retry schedule (timing & total retries) configurable

PROMS Server implementation

Testing

- An Entity Test Server was used for validating the methods
- Remote from pingback tester
- You can use it to →
- You can download the code and play!



Pingbacks Entity Server

This is a bare-bones RESTful web service, built to deliver HTML and RDF modules.

The register of datasets is available at <http://promsns.org/eg/entity-server/id/>

The, very simple, source code for this application is available publicly at <http://promsns.org/eg/entity-server/>

Contact the primary researcher involved, Nicholas Car (nicholas.car@csiro.au) if you have questions.

See <http://promsns.org/>

More Tools

- Online Message Validator →



Pingback message val

How to use the validator

1. formulate your message
 - you can use any system that can send HTTP requests like [cURL](#)
 - you can also use the purpose-built [PROMS Pingbacks message generator](#)
2. send the message, via an HTTP POST request, to the validator endpoint
3. review the validator results
 - valid messages will receive an HTTP status code response of 204, as per the [RFC](#)
 - error messages will be given with an appropriate HTTP status code (using the [HTTP status code](#))

Validator Endpoint

The validator endpoint is:

- <http://promsns.org/pingbacks/validator/>

More Tools

- Online Message Validator
- Pingbacks message generator →

Source

 master ▾ ...

PROMS pingbacks message generator


 [.gitignore](#)

 [functions.py](#)

 [LICENCE](#)

 [main.py](#)

 [README.md](#)

 [README.md](#)

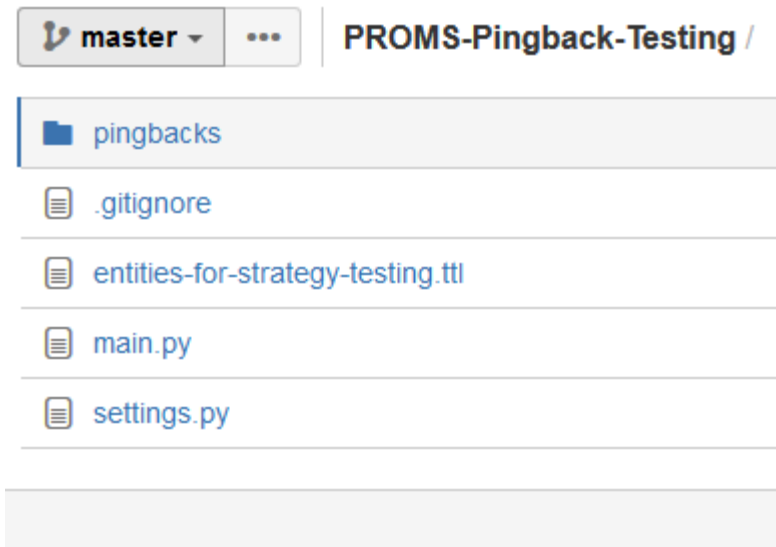
PROMS Pingbacks message

This small toolkit allows you to generate provenance pingbacks m
[PROMS wiki](#).

More Tools

- Online Message Validator
- Pingbacks message generator
- Whole Pingbacks module →

Source

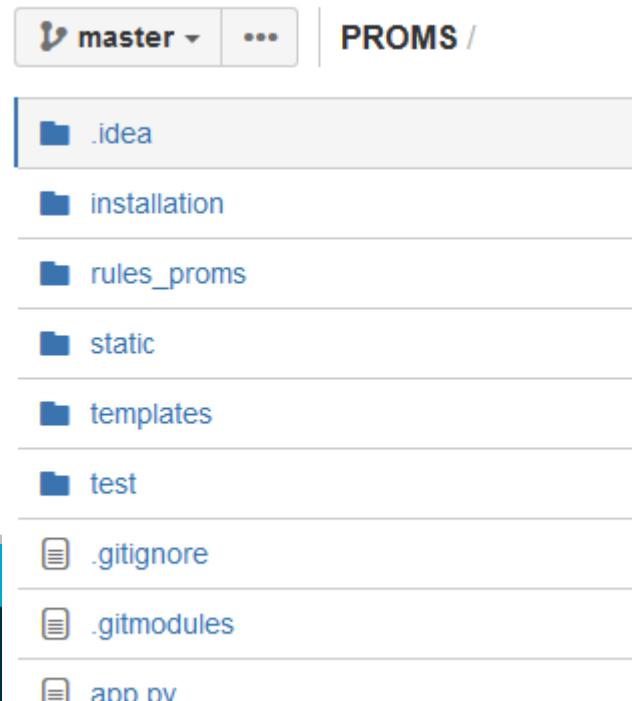


The screenshot shows a GitHub repository interface. At the top, there is a branch selector set to 'master' and a menu icon. To the right, the repository name 'PROMS-Pingback-Testing' is displayed. Below this, a list of files and folders is shown: a folder named 'pingbacks', and four files: '.gitignore', 'entities-for-strategy-testing.ttl', 'main.py', and 'settings.py'. Each item has a small icon representing its type (folder or file).

More Tools

- Online Message Validator
- Pingbacks message generator
- Whole Pingbacks module
- PROMS 3.1 →

Source

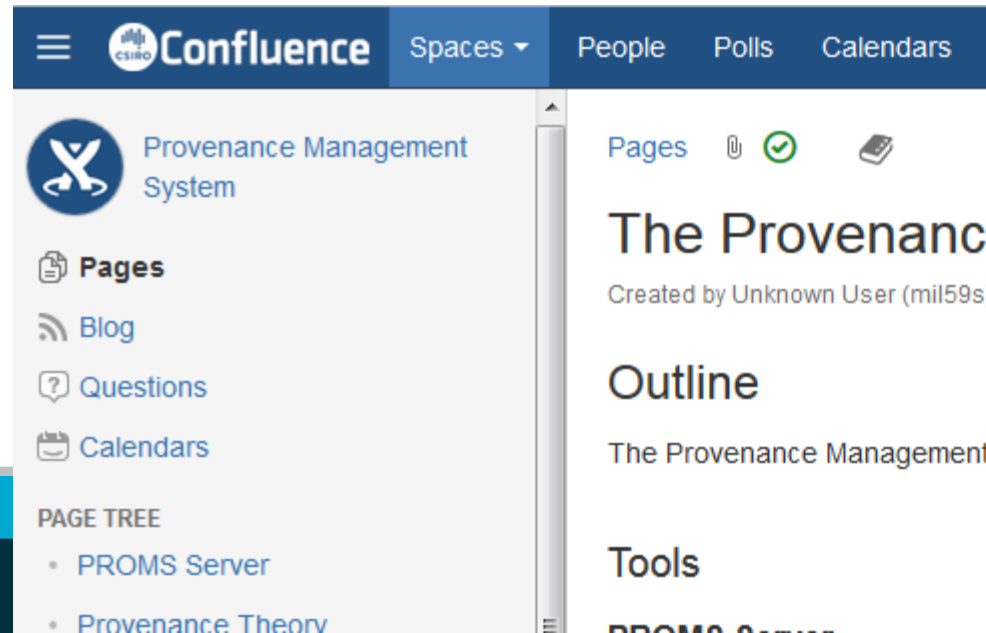


A screenshot of a GitHub repository interface for the 'PROMS' project. The repository is on the 'master' branch. The file list includes:

- .idea
- installation
- rules_proms
- static
- templates
- test
- .gitignore
- .gitmodules
- app.py

More Tools

- Online Message Validator
- Pingbacks message generator
- Whole Pingbacks module
- PROMS 3.1
- PROMS wiki (updated soon) →



The screenshot shows a Confluence page for the 'Provenance Management System'. The page header includes the Confluence logo, 'Spaces', 'People', 'Polls', and 'Calendars'. The main content area features a sidebar with navigation options: 'Pages', 'Blog', 'Questions', and 'Calendars'. Below the sidebar, there is a 'PAGE TREE' section with a list of items: 'PROMS Server' and 'Provenance Theory'. The main content area displays the title 'The Provenance Management System', the creator 'Created by Unknown User (mil59s)', and an 'Outline' section with the text 'The Provenance Management System'. Below the outline, there is a 'Tools' section with the text 'PROMS Server'.

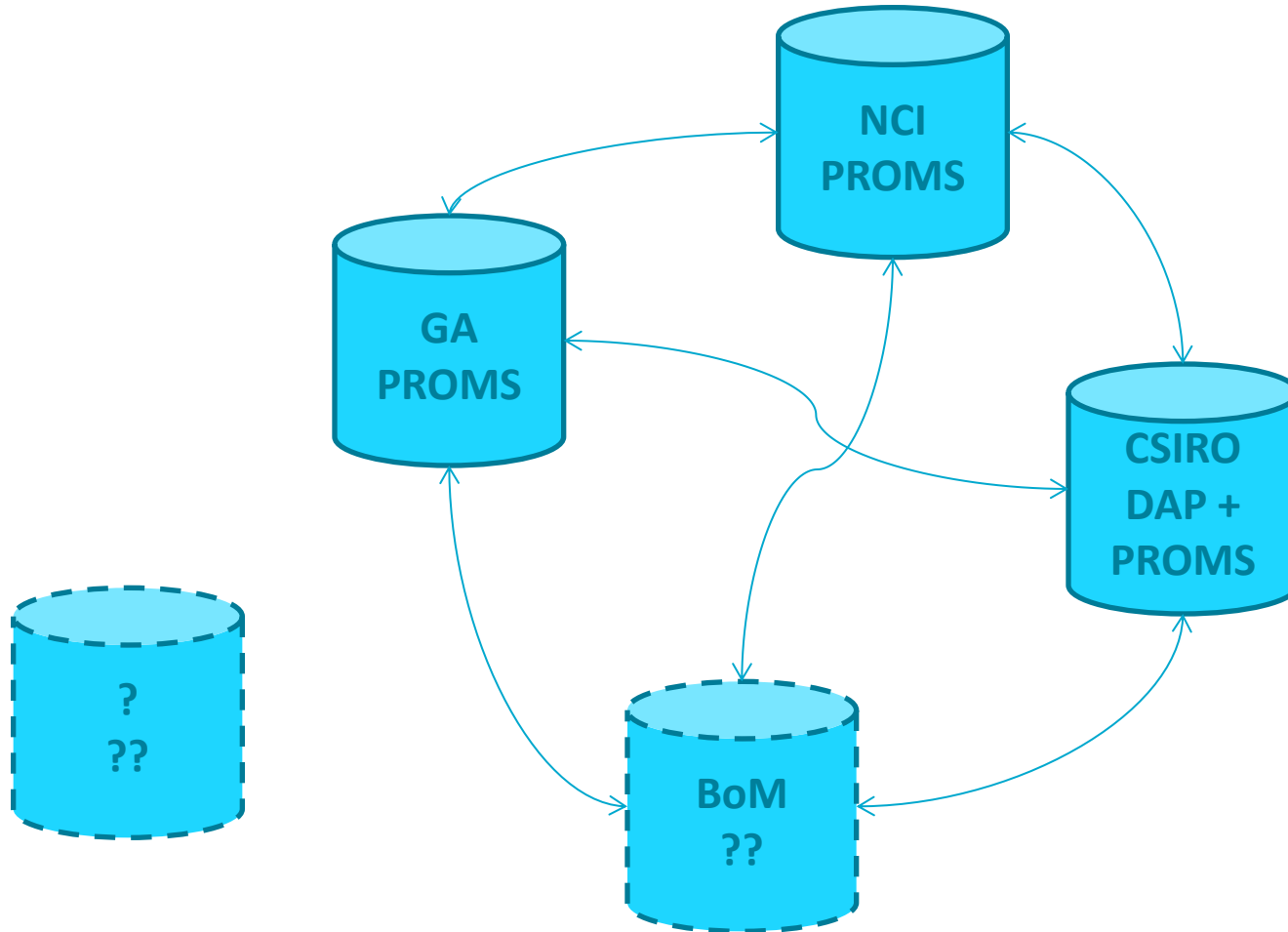
More Tools

- Online Message Validator
- Pingbacks message generator
- Whole Pingbacks module
- PROMS 3.1
- PROMS wiki (updated soon)
- Journal article (soon)

More Tools

- Online Message Validator
- Pingbacks message generator
- Whole Pingbacks module
- PROMS 3.1
- PROMS wiki (updated soon)
- Journal article (soon)
- All at <http://promsns.org>

An Australian Provenance Sharing Network



Cheers

Geoscience Australia
(formerly of CSIRO Land & Water)

Nicholas Car

t +61 2 6249 9093
e nicholas.car@ga.gov.au
w <http://promsns.org>

CSIRO LAND AND WATER
www.csiro.au

