An Overview on
PROV-AQ: Provenance Access and Query

Editors:
Graham Klyne
Paul Groth

Authors:
Luc Moreau
Olaf Hartig
Yogesh Simmhan
James Myers
Timothy Lebo
Khalid Belhajjame
Simon Miles
Stian Soiland-Reyes
Purpose

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

- Guiding principle:
  - Do not reinvent the wheel
  - Use standard Web protocols
Main Contributions

- Simple mechanisms for discovery and retrieval
- Provenance query mechanisms
  - Direct HTTP query service
  - (SPARQL query service)
- Provenance ping-back mechanism
Outline

Key Concepts
Simple Mechanisms
Provenance Query Services
Provenance Ping-back
Key Concepts

- Resource*
- Provenance record
- Provenance URI
- Target URI

Questions to be Addressed

- What is the provenance URI for a provenance record about a particular resource?
- What is the relevant target URI used in some provenance record?
- What if there is no provenance URI?
- What if there is no target URI?

Out of scope:
- Representation of provenance information
- Management of provenance information
Outline

Key Concepts

Simple Mechanisms

Provenance Query Services

Provenance Ping-back
Provenance URI and Target URI...

... for resources accessed by HTTP

- **Link header field** in the HTTP response

```
HTTP/1.1 200 OK
Content-Type: text/html
Date: Fri, 24 Feb 2012 12:58:02 GMT
Expires: Fri, 02 Mar 2012 12:57:52 GMT
Link: http://example.org/prov/aboutX; rel="http://www.w3.org/ns/prov#has_provenance"; anchor="http://example.org/X"
Content-Length: 18556

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
  <link rel="stylesheet" type="text/css" href="main.css" />
</head>
```

*defined in RFC 5988*
An Overview on PROV-AQ

Provenance URI and Target URI ...

... for resources accessed by HTTP

- Link header field* in the HTTP response

GET http://example.org/X HTTP/1.1
Accept: text/html

HTTP/1.1 200 OK
Content-Type: text/html
Date: Fri, 24 Feb 2012 12:58:02 GMT
Expires: Fri, 02 Mar 2012 12:57:52 GMT
Link: http://example.org/prov/aboutX; rel="http://www.w3.org/ns/prov#has_provenance";
anchor="http://example.org/X"

Content-Length:

*defined in RFC 5988
Provenance URI and Target URI ...

... for resources accessed by HTTP

- **Link header field*** in the HTTP response

GET http://example.org/X HTTP/1.1
Accept: text/html

HTTP/1.1 200 OK
Content-Type: text/html
Date: Fri, 24 Feb 2012 12:58:02 GMT
Expires: Fri, 02 Mar 2012 12:57:52 GMT

Link: http://example.org/prov/aboutX; rel="http://www.w3.org/ns/prov#has_provenance"; anchor="http://example.org/X"

Provenance URI

Target URI (optional)

*defined in RFC 5988
Embedding Provenance Links

- Links to provenance records may also be embedded in representations of a resource.
- Common mechanism for arbitrary representations is out of scope.
- PROV-AQ introduces mechanisms for:
  - HTML
  - RDF
Resources Represented as HTML

- `<link>` elements in the `<head>` section

```html
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <link rel="http://www.w3.org/ns/prov#has_provenance"
      href="http://example.org/prov/aboutX" />
    <link rel="http://www.w3.org/ns/prov#has_anchor"
      href="http://example.org/X" />
    <link rel="stylesheet" type="text/css"
      href="main.css" />
    <title>Welcome</title>
  </head>
  <body>
    <h1>Welcome to this Web page</h1>
  </body>
</html>
```
An Overview on PROV-AQ

Resources Represented as RDF

@prefix prov: <http://www.w3.org/ns/prov#>
<> dcterms:title "Welcome to example.org" ;
prov:has_anchor <http://example.org/X> ;
prov:has_provenance
    <http://example.org/prov/aboutX> .

RDF data
Outline

Key Concepts

Simple Mechanisms

Provenance Query Services

Provenance Ping-back
Provenance Query Services

- Provide provenance access mechanisms:
  - Simple HTTP-based query protocol for retrieving provenance records
  - SPARQL protocol for executing SPARQL queries over provenance records

- General procedure:
  1. Obtain service description
  2. Extract information about recognized query mechanisms and how to use them
  3. Select mechanism and use it as described
Obtain a Service Description

- Dereference the service-URI
- Resource providers may indicate service-URIs using the relation type denoted by:
  \[ \text{http://www.w3.org/ns/prov#has_query_service} \]
  - e.g., via a Link header field in an HTTP response:

  \[
  \text{Link: service-URI; rel="http://www.w3.org/ns/prov#has_query_service"; anchor="target-URI"}
  \]

  - e.g., via an RDF triple in an RDF representation:

  \[
  <> \text{dcterms:title "Welcome to example.org" ; prov:has_query_service <service-URI>}
  \]
Service Description (in RDF)

@prefix prov: <http://www.w3.org/ns/prov#>
@prefix sd: <http://www.w3.org/ns/sparql-service-description#>

<service-URI> a prov:ServiceDescription ;
    prov:describesService <sparql-query-service> ;
    prov:describesService <direct-query-service> .

<sparql-query-service> a sd:Service ;
    sd:endpoint <sparql-endpoint-URI> .

<direct-query-service> a prov:DirectQueryService ;
    prov:provenanceUriTemplate "direct-query-template" .
Using a Direct Query Service

- For retrieving provenance about a target-URI (that needs to be specified via a query URI)
- Possible query URIs described by **URI template**
  - Indicated by `prov:provenanceUriTemplate`
  - Variable `uri` in the template stands for the target-URI
  - Example template: `http://example.org/pquery?target={uri}`
  - Example invocation:

```bash
GET /pquery?target=http%3A%2F%2Fexample.org/X HTTP/1.1
Host: example.org
```

*defined in RFC 3986*
SPARQL Query Service

- For use cases where aforementioned mechanisms are insufficient
- Example: target-URI unknown

```
SELECT ?provenance_uri WHERE {
  prov:has_provenance ?provenance_uri . }
```

- Example: ask for a particular piece of provenance information only

```
SELECT ?start_time WHERE {
  <http://example.org/X> prov:wasGeneratedBy ?act .
  ?act prov:startedAtTime ?start_time .}
```
Outline

Key Concepts
Simple Mechanisms
Provenance Query Services
Provenance Ping-back
Purpose

- For discovering related provenance information that resource publishers would not know about
  - e.g., how a resource is used by third parties
- Ping-back enables cooperation among resource publishers and using parties
  - Provenance information about the use of a resource can be posted to the publisher
Provenance Ping-Back URI

- Indicated using the relation type denoted by:
  
  \[\text{http://www.w3.org/ns/prov#pingback}\]

- e.g., via a Link header field in an HTTP response:
  
  \[
  \text{Link: ping-back-URI; rel="http://www.w3.org/ns/prov#pingback"}
  \]

- Users may post to the ping-back URI a list of provenance-URIs (denoting provenance records that describe uses of the resource)

  \[
  \text{POST http://example.org/X/provping HTTP/1.1}
  \text{Content-Type: text/uri-list}
  \text{http://acme.com/Y/provenance}
  \text{http://acme.com/another/provenance}
  \]
Summary

- Simple mechanisms for discovery and retrieval
- Provenance query mechanisms
  - Direct HTTP query service
  - (SPARQL query service)
- Provenance ping-back mechanism