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

Presentation at the Dagstuhl Seminar on Principles of Provenance
February 29, 2012
Purpose

- Describes how to
  - locate,
  - retrieve, and
  - query
  provenance information

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

- Simple mechanisms for discovery and retrieval
- More advanced
  - discovery service and
  - query mechanisms
Outline

Key Concepts
Simple Mechanisms
Provenance Services
Querying Provenance Information
Open Issues
Key Concepts

- Resource*
- Entity
- Provenance information
- Provenance URI
- Entity URI

Questions to be Addressed

- What is the provenance URI for provenance information about a particular entity?
- What is the relevant entity URI used in some provenance information?
- What if there is no provenance URI?
- What if there is no entity URI?

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

Key Concepts

Simple Mechanisms

Provenance Services

Querying Provenance Information

Open Issues
An Overview on PROV-AQ

... for resources accessed by HTTP

- Link header field* in the HTTP response

```
GET //example.org/X HTTP/1.1
Host: example.org
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="provenance"; anchor="http://example.org/X"
Content-Length: 18556
```

*defined in RFC 5988

<?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" />
    <!-- page-specific style definitions -->
    <style type="text/css">
```
Entity URI and Provenance URI...

... for resources accessed by HTTP

- Link header field* in the HTTP response

GET //example.org/X HTTP/1.1
Host: example.org
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="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" />
<!-- page-specific style definitions -->
<style type="text/css">

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

... for resources accessed by HTTP

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

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

Link: http://example.org/prov/aboutX; rel="provenance"; anchor="http://example.org/X"
Content-Length: 18556

*defined in RFC 5988
Embedding Provenance Links ...

... in representations of resources.

- Proposal for particular representation formats:
  - HTML
  - RDF based (e.g. RDF/XML, Turtle, RDFa)

- Common mechanism for arbitrary representations is out of scope
Outline

Key Concepts
Simple Mechanisms
Provenance Services
Querying Provenance Information
Open Issues
Functionality

- Provenance retrieval service
  - Input: entity URI
  - Output: provenance information

- Provenance discovery service
  - Input: entity URI
  - Output: provenance URI(s)

- Provided via a REST API
Service Discovery

- Simple mechanism
  - ... for resources accessed by HTTP:
    Link: Service-URI; rel="provenance-service"; anchor="Entity-URI"

- However, a general approach is out of scope
1. Service description resource
   - Accessible by dereferencing the Service-URI
   - How to construct URIs for other two resource types
2. Provenance information resource
   - Provides provenance information
   - Represented in any format (e.g. RDF based, XML)
3. Provenance locations resource
Provenance Locations Resource

- Enumerates provenance URIs for a particular entity
Provenance Locations Resource

- Enumerates provenance URIs for a particular entity
- Example representation in JSON format

```json
{  "uri": "http://example.org/X",
   "provenance": [
      "http://prov.example.org/aboutX",
      "http://example.com/prov?id=X"
   ]
}
```
Provenance Locations Resource

- Enumerates provenance URIs for a particular entity
- Example representation in JSON format

```json
{ "uri": "http://example.org/X",
  "provenance": [
    "http://prov.example.org/aboutX",
    "http://example.com/prov?id=X"
  ]
}
```

- Example representation in Turtle (i.e. RDF data)

```turtle
@prefix prov: <@@TBD@@#> .
<http://example.org/X> a prov:Entity ;
  prov:hasProvenance <http://prov.example.org/aboutX> ;
  prov:hasProvenance <http://example.com/prov?id=X> .
```
Using the API ...

... to retrieve provenance URI(s)

- Required: *Entity-URI and Service-URI*

1. Obtain service description
   (by dereferencing *Service-URI*)

2. Extract URI template for locations resources

3. Form *Provenance-Locations-URI*
   (by instantiating the template with *Entity-URI*)

4. Obtain provenance locations resource
   (by dereferencing *Provenance-Locations-URI*)

5. Extract relevant *Provenance-URIs*
Outline

Key Concepts

Simple Mechanisms

Provenance Services

Querying Provenance Information

Open Issues
Provenance Query Services

- Alternative for cases where aforementioned mechanisms are insufficient
- PROV-AQ recommends SPARQL
  - Query language for RDF data
  - Protocol for SPARQL query processing services
Example Queries

- **Use case: unknown entity URI**

```
SELECT ?provenance_uri WHERE {
  ?entity prism:doi "1234.5678" ;
  prov:hasProvenance ?provenance_uri . }
```

- **Use case: ask for a particular piece of provenance information only**

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

Key Concepts

Simple Mechanisms

Provenance Services

Querying Provenance Information

Open Issues
Main Open Issues

- Entities disjoint from Activities in PROV-DM
- Remove service descriptions (i.e. prescribe a particular template)?

```json
{
  "provenance_service_uri": "http://example.org/PS/",
  "provenance_locations_template": "http://ps.example.org/loc?uri={uri}"}
```

- Simplifies the process of using provenance services
- Embedding provenance information directly
  - Best practice recommendation for HTML: RDFa
  - General approach is out of scope
Main contributions:

- Simple mechanisms for retrieving and discovering provenance information
- More advanced discovery service and query mechanisms
Backup Slides
Resources Represented as HTML

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

```html
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
  <link rel="provenance" href="http://example.org/prov/aboutX" />
  <link rel="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>
  <p>
```
Provenance Service URI ...

... for resources accessed by HTTP

Link: Service-URI; rel="provenance-service"; anchor="Entity-URI"
Resources Represented as HTML

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

```html
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <link rel="provenance" href="http://example.org/prov0fX" />
    <link rel="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>
    <p>Indicating a provenance service:
      <link rel="provenance-service" href="Service-URI" />
    </p>
  </body>
</html>
```
Provides templates for URIs of

- provenance locations resources and/or
- provenance information resources
Service Description

- Provides templates for URIs of
  - provenance locations resources and/or
  - provenance information resources
- Example description represented in JSON

```json
{
  "provenance_service_uri": "http://example.org/PS/",
  "provenance_locations_template": "http://ps.example.org/loc?uri={uri}",
  "provenance_content_template": "http://example.org/PS/prov?uri={uri}"
}
```
Service Description

- Provides templates for URIs of
  - provenance locations resources and/or
  - provenance information resources
- Example description represented in JSON

```json
{
  "provenance_service_uri": "http://example.org/PS/",
  "provenance_locations_template":
    "http://ps.example.org/loc?uri={uri}",
  "provenance_content_template":
    "http://example.org/PS/prov?uri={uri}"  
}
```

- Provenance locations resource for
  http://example.org has URI:

http://ps.example.org/loc?uri=http%3A%2F%2Fexample.org
Using the API ...

... to retrieve provenance information

- Required: *Entity-URI* and *Service-URI*

1. Obtain service description
   (by dereferencing *Service-URI*)

2. Extract URI template for information resources

3. Form *Provenance-URI*
   (by instantiating the template with *Entity-URI*)

4. Obtain provenance information resource
   (by dereferencing *Provenance-URI*)

5. Extract relevant provenance information