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 Semantic Data Management
April 24, 2012
Purpose

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

- Guiding principle:
  - Do not reinvent the wheel
  - Use standard Web protocols
Key Concepts

- Resource*
- Provenance information
- Provenance URI
- Target URI

Questions to be Addressed

- What is the *provenance URI* for *provenance information* about a particular *resource*?
- What is the relevant *target URI* used in some *provenance information*?
- 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
Main Contributions

- Simple mechanisms for discovery and retrieval
- Retrieval service
- Query service
Main Contributions

- Simple mechanisms for discovery and retrieval
  - Retrieval service
  - Query service
An Overview on PROV-AQ

... for resources accessed by HTTP

- Link header field* in the HTTP response

```http
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
<?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
```
... 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
An Overview on PROV-AQ

Provenance URI and Target URI...

... for resources accessed by HTTP

- Link header field* in the HTTP response

```plaintext
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">

Provenance URI

Target URI (optional)

*defined in RFC 5988
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>
```
Other Formats

- Resources represented as RDF

```rdfs
@prefix prov: <http://www.w3.org/ns/prov#>;
<> dcterms:title "Welcome to example.org" ;
prov:hasAnchor <http://example.org/X> ;
prov:hasProvenance <http://example.org/prov/aboutX> .
```

- Common mechanism for arbitrary representations is out of scope
Main Contributions

- Simple mechanisms for discovery and retrieval
- Retrieval service
- Query service
Functionality

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

- HTTP GET request for service URI combined with target URI:

```
GET /prov/service?target=http://example.org/X HTTP/1.1
Host: example.org
```

Service URI
http://example.org/prov/service

Target URI
http://example.org/X
Simple mechanism

... for resources accessed by HTTP:

```
Link: Service-URI; rel="provenance-service";
    anchor="Target-URI"
```

... for resources represented as HTML:

```
<link rel="provenance-service"
    href="Service-URI" />
```

However, a general approach is out of scope
Main Contributions

- Simple mechanisms for discovery and retrieval
- Retrieval service
- Query service
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 *target URI*

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

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

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

- Main contributions:
  - Simple mechanisms for retrieving and discovering provenance information
  - Provenance retrieval service
  - Provenance query service

- Status:
  - First Working Draft, 10 January 2012
  - Next release in a couple of weeks
These slides have been created by Olaf Hartig
http://olafhartig.de

This work is licensed under a Creative Commons Attribution-Share Alike 3.0 License
http://creativecommons.org/licenses/by-sa/3.0/