ProvStore: A Public Provenance Repository

Trung Dong Huynh(✉) and Luc Moreau

Electronics and Computer Science, University of Southampton,
Southampton SO17 1BJ, UK
{tdh,l.moreau}@ecs.soton.ac.uk

Abstract. ProvStore is the first online public provenance repository supporting the new PROV standards by W3C. It allows users and applications to store and (optionally) publish the provenance of their data on the Web. Provenance documents can be transformed, visualized, and shared in various serializations, with all the functionality also available to third-party applications via a RESTful API (OAuth supported).

1 Provenance Repository

ProvStore (https://provenance.ecs.soton.ac.uk/store/) is the first public repository of provenance documents supporting the PROV standards for provenance on the Web by the World Wide Web Consortium [MM13]. Users can register for a free account, allowing them to upload and share provenance documents either privately or publicly in various representations (see Fig. 1 for an example[1]). Specifically, it supports the Provenance Notation (PROV-N), RDF encoded using the PROV Ontology (PROV-O) in Turtle or TriG formats, PROV-XML, and PROV-JSON [HJK+13].

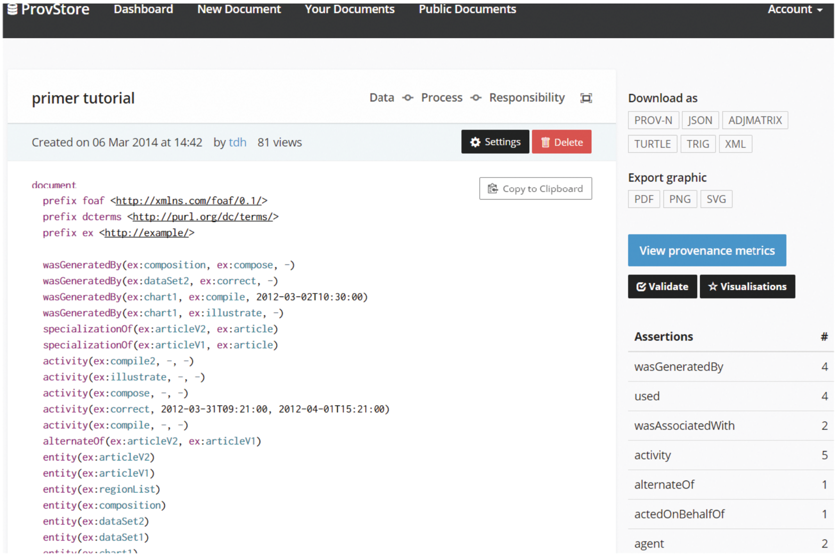
By default, documents submitted to ProvStore are private and can only be accessed by their owners. Document owners, however, can choose to share their documents with others in two ways: making a document *public*, i.e. available to any visitor to ProvStore, or sharing it with specific ProvStore's users. The former is useful for users who want to expose the provenance of their resources (e.g. papers, reports, data sets) to the public; the link to a document on ProvStore can be attached as the provenance URI along with the corresponding resource.[2] In the latter, different access roles can be set to authorized users for fine-grain access control: administrator, editor, contributor, or reader. Except reader, all other roles and the owner can append new provenance bundles to a document after it has been created. It is suitable for sharing provenance between a team of collaborating humans and/or applications (see Sect. 3 for more information about the application programming interface provided by ProvStore).

ProvStore was funded by the UK Engineering and Physical Sciences Research Council (EPSRC) as part of project Orchid, grant EP/I011587/1.

[1] Online address: https://provenance.ecs.soton.ac.uk/store/documents/1979/.

[2] See www.w3.org/TR/prov-aq for more information on provenance access and query. Document links on ProvStore support HTTP content negotiation. For example, if the HTTP request specify a header `Accept: application/json`, the PROV-JSON representation of the provenance document will be returned.

B. Ludäscher and B. Plale (Eds.): IPAW 2014, LNCS 8628, pp. 275–277, 2015.
DOI: 10.1007/978-3-319-16462-5_32



primer tutorial

Created on 06 Mar 2014 at 14:42 by tdh 81 views

```
document
  prefix foaf <http://xmlns.com/foaf/0.1/>
  prefix dcterms <http://purl.org/dc/terms/>
  prefix ex <http://example/>

  wasGeneratedBy(ex:composition, ex:compose, -)
  wasGeneratedBy(ex:dataSet2, ex:correct, -)
  wasGeneratedBy(ex:chart1, ex:compile, 2012-03-02T10:30:00)
  wasGeneratedBy(ex:chart1, ex:illustrate, -)
  specializationOf(ex:articleV2, ex:article)
  specializationOf(ex:articleV1, ex:article)
  activity(ex:compile2, -, -)
  activity(ex:illustrate, -, -)
  activity(ex:compose, -, -)
  activity(ex:correct, 2012-03-31T09:21:00, 2012-04-01T15:21:00)
  activity(ex:compile, -, -)
  alternateOf(ex:articleV2, ex:articleV1)
  entity(ex:articleV2)
  entity(ex:articleV1)
  entity(ex:regionList)
  entity(ex:composition)
  entity(ex:dataSet2)
  entity(ex:dataSet1)
  entity(ex:chart1)
```

Assertions	#
wasGeneratedBy	4
used	4
wasAssociatedWith	2
activity	5
alternateOf	1
actedOnBehalfOf	1
agent	2

Fig. 1. The screen-shot of a ProvStore document.

On each document (Fig. 1), users can see its provenance descriptions in PROV-N, along with some statistics about the numbers of assertions. ProvStore also provides a number of provenance network metrics [EHM+12] calculated on the graph representation of the document. As mentioned above, access links to various provenance representations are included, in addition to a numbers of provenance transformations and visualizations (see Sect. 2). The provenance of the document can be checked directly from inside the document page (provided by the external ProvValidator service[3]).

2 Provenance Transformation and Visualization

A provenance document can contain bundles, which are a PROV construct to support bundling a set of provenance descriptions (so allowing provenance of provenance to be expressed) [MM13]. To support relating provenance statements within a document across its bundles, ProvStore can produce a *flattened* representation of the document in which all of its provenance statements are merged into a flat document. In this representation, the provenance of entities distributed in multiple bundles can be "connected" for further examination.

In addition to the flattened representation, ProvStore provides a number of provenance views: Data Flow (concerned with the flow of information or the transformations of things), Process Flow (concerned with the processes that took place), and Responsibility (assigning responsibility for what happened) [MG13, Chap. 3].

[3] provenance.ecs.soton.ac.uk/validator.

These views are simplified versions of the original document produced by selecting only the relevant provenance descriptions from it. They can facilitate the examination of provenance information by allowing users to focus on a single aspect of it rather than the full descriptions. Each of the views can be applied either on the original document or its flattened version.

All versions (original or flattened, optionally simplified in a provenance view) of a ProvStore document can be visualized in a (static) graphical representation (in the SVG, PNG, or PDF formats). In addition, ProvStore provides interactive visualization tools for users to explore a provenance graph through a Hive plot (highlighting input, output, and intermediary nodes), a Wheel plot (showing the density of connections to/from nodes), a Gantt chart (presenting entities, activities, and agents on a time line), and a Sankey diagram (showing flows of 'influence' between provenance elements). All the interactive visualizations, except the Gantt chart, also allow filtering on provenance assertion types to simplify the visualizations.

3 RESTful Application Programming Interface (API)

All of the functionality described in the previous sections (with the exception of interactive features like validation and visualizations) can be accessed programmatically via a RESTful API[4] over the Hypertext Transfer Protocol. ProvStore, hence, can serve as a provenance storage-and-publish service on the cloud, providing applications a means to make the provenance of their data available online as soon as it is generated/recorded. Authorized applications must authenticate with ProvStore's API either by using their (revocable) secret API keys or by following the OAuth (version 1) protocol. With the latter, ProvStore enables users of any third-party applications or web sites (that registered with it) to store or access their provenance data directly from inside such applications in a seamless fashion.

References

[EHM+12] Ebden, M., Huynh, T.D., Moreau, L., Ramchurn, S., Roberts, S.: Network analysis on provenance graphs from a crowdsourcing application. In: Groth, P., Frew, J. (eds.) IPAW 2012. LNCS, vol. 7525, pp. 168–182. Springer, Heidelberg (2012)

[HJK+13] Huynh, T.D., Jewell, M.O., Sezavar Keshavarz, A., Michaelides, D.T., Yang, H., Moreau, L.: The PROV-JSON serialization. Technical report, World Wide Web Consortium, April 2013

[MG13] Moreau, L., Groth, P.: Provenance: An Introduction to PROV. Morgan & Claypool, San Rafael (2013)

[MM13] Moreau, L., Missier, P.: PROV-DM: The PROV Data Model. Technical report, World Wide Web Consortium, W3C Recommendation (2013)

[4] See provenance.ecs.soton.ac.uk/store/help/api for the full specification of the API and example codes.