

# PREMIS Event Service

Mark Phillips  
University of North Texas  
Denton  
TX 76205, USA  
+1 (940) 565-2415

Mark.Phillips@unt.edu

Matt Schultz  
Educopia Institute  
Atlanta  
GA 30309, USA  
+1 (616) 566-3204

Matt.Schultz@metaarchive.org

Kurt Nordstrom  
University of North Texas  
Denton  
TX 76205, USA  
+1 (940) 369-7809

Kurt.Nordstrom@unt.edu

## ABSTRACT

The University of North Texas Libraries has created a digital library infrastructure that is designed using the Curation Micro Services methodology for building repositories. Based on modular components designed for reuse and re-configuration the authors propose a general-purpose preservation event logging system using the PREMIS Event and Agent data model. An example implementation developed with the Django Web framework and employing RESTful design patterns with the Atom Publishing Protocol is described.

## Categories and Subject Descriptors

H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia – *Architectures, Navigation.*

## General Terms

Curation Micro Services

## Keywords

Digital Preservation, Repositories, PREMIS, Curation Micro Services

## 1. INTRODUCTION

The University of North Texas (UNT) has implemented a robust architecture for digital library initiatives utilizing the Curation Micro Services methodology for building repository infrastructure. Consisting of various loosely coupled and highly integrated components, this architectural design has provided a new set of abstractions for common repository functions to be developed and implemented in the production system. One micro-service that has been identified as a generally useful and reusable component in this infrastructure is an event logging service based on the PREMIS Event data model and utilizing Web architecture practices such as REST and the Atom Publishing Protocol.

## 2. BACKGROUND WORK

The UNT Libraries has been involved in the building of digital library services for a number of years and in the summer of 2009 released a new infrastructure based on the Curation Micro-Services methodology pioneered by the California Digital Libraries' UC3 group. [1] This architectural model emphasizes single purpose, task based components that are loosely coupled and highly integrated into a system's overall design. This design methodology allows for the independent creation and deployment of services that provide various functions to a repository infrastructure. These services can be changed over time as new technologies are introduced or if a service no longer fulfills a

logical role in the repository design. In any digital preservation repository there is a need to create and store metadata related to events that occur in the lifecycle of the digital objects that are ingested and curated in the system. The UNT Libraries has developed a reusable service that can be used as a central event logging service for a repository.

Based on the PREMIS model for both events and agents, the PREMIS Event Service is a Django Web application designed using the Atom Publishing Protocol in a RESTful architecture. [2] [3] [4] [5] This event service is currently in production at the UNT Libraries in the underlying repository infrastructure for the UNT Digital Library and the Portal to Texas History. [6] [7] Examples of events that are currently logged with the system include ingest, replication, fixity, and deletion events. As new repository services are placed in production the record of their interaction with the digital objects in the system will be added to the event services. Examples of future services include virus checking, content normalization, and transformation events.

## 3. SYSTEM DESCRIPTION

The PREMIS Event Service utilizes the Event and Agent data model from the PREMIS Data Dictionary as the underlying model for events being logged. This model was chosen because of its general utility in the area of digital preservation as well as the growing acceptance of PREMIS in the community. The Python-based Web framework Django is used to provide a Web accessible interface for the service. The REST design pattern was used in conjunction with the Atom Publishing Protocol (AtomPub) to provide a standard way of creating and accessing events in the logging system.

In a given workflow, for a particular service defined within an repository infrastructure, the final step in its processing sequence will be to define and send a PREMIS Event entry into the event logging service, detailing the outcome of the given operation. Typically, the output falls into the category of "pass" or "failure", though the data model provides for much more specific reporting. Within the scope of the UNT leverage of the service, most of the data values chosen are part of controlled vocabularies that provide actionable URLs for each term, giving canonical meaning for each specific data value. While the PREMIS specification does not enforce this practice, it is encouraged.

Events are added and retrieved using AtomPub conventions. Single events are contained within ATOM <entry> tags, and multiple events are returned as an ATOM <feed>. Keeping with the philosophy of REST, the standard HTTP operations of POST and GET are the basis for item creation and retrieval.

Upon event creation, the PREMIS XML is parsed and stored in an internal data model that can be aggregated and queried against. This allows for sophisticated retrieval of specific events, based upon various criteria. Parameters that can be specified during retrieval include (but are not limited to) the date of the event, the PREMIS Agent responsible for the event, the type of event, the outcome of the event or a specific object identifier that the event acted upon. Parameters can be combined to further refine retrieval sets. These parameters are passed as part of the feed retrieval URL.

Finally, the PREMIS Event Service provides a “human consumable” web interface that exists in parallel with the AtomPub implementation. This interface offers a user-friendly console, search fields and formatted output of data, much like using any standard search engine. The same granularity of queries and retrieval exists, but the output is returned in tabular form, as opposed to XML. This interface is intended to facilitate overall reporting and monitoring by repository managers. The URLs are chosen to easily and intuitively facilitate switching between the AtomPub and human-readable views of the data.

The PREMIS Event Service has been in production since November 2010 and has provided the necessary functionality in the UNT Libraries digital library systems as expected. Future enhancements include aggregating and reporting statistics based on logged events, which will help administrators in the ongoing management of their repositories.

#### **4. CONCLUSION**

As the interest in building repositories using the Curation Micro-Services approach continues to grow it is important to define

modular and reusable services that can be shared and used in a variety of configurations in different environments. Repositories will always have different scope, mission and users but there tend to be a set of standard mechanisms and processes, which can leverage common tools and approaches. The PREMIS Event Service is an example of such a service that is usable in a variety of situations and can be leveraged in many repository environments to provide the necessary functions of preservation event logging for a repository infrastructure.

#### **5. ACKNOWLEDGMENTS**

This research was funded in part by the National Historic Publications and Records Commission (NHPRC).

#### **6. REFERENCES**

- [1] Curation Micro Services - <http://www.cdlib.org/services/uc3/curation/>
- [2] PREMIS - <http://www.loc.gov/standards/premis/>
- [3] REST - <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>
- [4] Atom Publishing Protocol - <http://tools.ietf.org/html/rfc5023>
- [5] Django - <http://www.djangoproject.com/>
- [6] UNT Digital Library – <http://digital.library.unt.edu>
- [7] Portal to Texas History – <http://texashistory.unt.edu>