

# Memento

## Uniform and Robust Access to Resource Versions

Herbert Van de Sompel  
Los Alamos National Laboratory  
PO Box 1663  
Los Alamos, NM, USA  
herbertv@lanl.gov

### ABSTRACT

The Memento protocol tightly integrates the Web of the Present and that of the Past, making it possible to seamlessly navigate between both. The protocol defines an interoperable approach to access versions of a resource in web archives or content management systems such as wikis that leverage the URI of that resource and the datetime of the required resource version. Technically, the Memento protocol is an extension of HTTP that is fully based on the primitives of Web interoperability: URIs, resource representations, links, content negotiation. The tutorial will give an in-depth insight in various aspects of the Memento protocol that meanwhile has been published as RFC 7089.

### General Terms

Infrastructure, preservation strategies and workflows, studies and best practice

### Keywords

Versioning, web archives, content management systems, HTTP, content negotiation, interoperability, web persistence, internet robustness

## 1. TUTORIAL OUTLINE

The tutorial will provide a detailed insight in various aspects of the Memento “Time Travel for the Web” protocol. The tutorial is aimed to be useful for developers interested in implementing Memento compliant clients or servers, and project managers, information architects, repository administrators interested in learning whether and how Memento concepts can be used to meet challenges they face in the realm of resource versioning.

The remainder of this section details the focus areas of the tutorial.

### 1.1 Motivation

The tutorial will start by providing an insight in the motivation for the multi-year Memento effort, which is to be found in the poor integration between the Present and the Past Web. This lack of integration is exemplified by problems related to navigating from the current version of a resource to past versions, from past versions to the current version, and to consistently navigate the Web of the Past.

### 1.2 Memento Protocol

iPres 2014 conference proceedings will be made available under a Creative Commons license.

With the exception of any logos, emblems, trademarks or other nominated third-party images/text, this work is available for reuse under a Creative Commons Attribution 3.0 unported license. Authorship of this work must be attributed. View a [copy of this licence](#).

The effort to specify the Memento protocol started in late 2009 and concluded in December 2013 with the publication of the specification as an IETF RFC [1]. The core ingredients of the protocol will be introduced (datetime negotiation, Original Resource, TimeGate, TimeMap, Memento, Memento HTTP Headers) and the client-server interactions will be detailed for various patterns that differ mainly in whether an Original Resource, its TimeGate, and its Mementos reside on the same server or not. Special attention will be given to aspects of Memento Aggregation, which allows locating the temporally most appropriate archived resource version across web archives.

### 1.3 Memento and Resource Versioning

The Memento protocol is closely aligned with a common resource versioning pattern that consists of:

- Having a generic URI where at any moment in time the current version of the resource is accessible.
- Having a dedicated version URI for each resource version.

Systems that support this resource versioning pattern do not necessarily need to implement the entire protocol at once but can gradually implement aspects of it in a modular manner, with each step along the path providing increased functionality regarding access to resource versions. The incremental steps, as described in [2], will be explained:

- Providing HTTP response headers for resource versions to convey version date and links
- Publishing a TimeMap, a list of resource versions
- Exposing a TimeGate that supports datetime negotiation to access resource versions

### 1.4 Memento and Web Persistence

Memento’s time travel capability provides an essential ingredient to address the well-known link rot, also known as “404 Not Found”, problem. If a link is broken, follow it back into the past and obtain a version from a web archive or resource versioning system. But, as described in [3], in order to fully tackle the problem, several open questions remain to be answered, including: Which date should be used for time travel; How to convey information about a known archival version of a linked resource in an HTML page, and how to make sure such archival versions are created in the first place? The tutorial will provide insights in the thinking of two ongoing activities with this regard:

- Hiberlink<sup>1</sup>, a Mellon-funded collaboration between the Los Alamos National Laboratory and the University of

---

<sup>1</sup> Hiberlink, <http://hiberlink.org>

Edinburgh that investigates the extent and nature of reference rot in web-based scholarly communication and explores approaches to ameliorate the problem. The project is inspired by a 2011 pilot study [4] that quantified scholarly link rot at an unprecedented scale.

- Internet Robustness<sup>2</sup>, a collaboration between Harvard University, the Los Alamos National Laboratory, and Old Dominion University aimed at increasing link robustness by specifying how to express information about archival versions of resources that are linked from an HTML page. The project has close ties with the study pertaining to reference rot in legal citations [5] and the perma.cc<sup>3</sup> effort aimed at pro-actively archiving resources linked from legal literature.

## 1.5 Memento Tools

A wide range of Memento compliant tools is meanwhile available, and the tutorial will provide an overview of the most prominent server-side and client-side ones, including Global Open Wayback, SiteStory, Memento MediaWiki extension, Memento Time Travel for Chrome, and mcurl.

## 1.6 Memento at Work

The power of Memento's time travel will be illustrated by means of demonstration of both production and experimental versions of Memento-related tools.

## 2. ACKNOWLEDGMENTS

During its initial years, the Memento project received funding from the Library of Congress. Hiberlink is funded by the Andrew W. Mellon Foundation.

## 3. REFERENCES

- [1] Van de Sompel, H., Nelson, M. L., and Sanderson, R. 2013. RFC 7089: HTTP Framework for Time-Based Access to Resource States – Memento. <http://ietf.org/rfc/rfc7089.txt>
- [2] Van de Sompel, H. 2013. Memento Guide: Resource Versioning and Memento. <http://mementoweb.org/guide/howto/>
- [3] Van de Sompel, H., Klein, M., Sanderson, R., and Nelson, M.L. 2013. Thoughts on Referencing, Linking, Reference Rot. <http://mementoweb.org/missing-link/>
- [4] Sanderson, R., Phillips, M., and Van de Sompel, H. 2011. Analyzing the persistence of referenced web resources with Memento. <http://arxiv.org/abs/1105.3459>
- [5] Zittrain, J., Albert, K., and Lessig, L. 2014. Perma: Scoping and addressing the problem of link and reference rot in legal citations. [http://www.harvardlawreview.org/issues/127/february14/forum\\_1031.php](http://www.harvardlawreview.org/issues/127/february14/forum_1031.php)

---

<sup>2</sup> Internet Robustness, <http://cyber.law.harvard.edu/research/internetrobustness>

<sup>3</sup> perma.cc, <http://perma.cc>