Time Travel for the Web

Herbert Van de Sompel
Michael L. Nelson
Robert Sanderson
Lyudmila Balakireva
Harihar Shankar
Scott Ainsworth

http://mementoweb.org/
The Web has become a major carrier of our cultural record.

Hence, significant attention is being paid to safeguarding portions of the Web of the Past.
Memento wants to make it easy
to access the Web of the Past.
Memento achieves this by technically integrating the present Web and the past Web, by introducing a uniform version access capability for the Web.
Content Management Systems:

- Designed to be aware of all versions of a resource;
- Self-contained;
- Variety of proprietary version mechanisms;
- Versions interlinked using proprietary mechanisms.
World Wide Web:

- Designed to forget about prior versions of a resource;
- Distributed.

Resource state may evolve over time. Requiring a URI owner to publish a new URI for each change in resource state would lead to a significant number of broken references. For robustness, Web architecture promotes independence between an identifier and the state of the identified resource.
There are resource versions on the Web:

- Content Management Systems;
- Web Archives;
- Transactional archives;
- Search engine caches.
But the Web architecture has no way to deal with them:

• Cannot talk about a resource as it used to exist;

• Cannot access a prior version knowing the current one;

• Cannot access the current version knowing a prior one;

Current approaches are ad hoc and localized.
Memento:

- Looks at the Web as a Content Management System;
- Introduces the uniform capability to access versions on the Web;
- Does not build new archives but leverages all systems that host versions: Web archives, Content Management Systems, Software Version Systems, etc.
Memento’s version access approach:

- Is distributed: versions may exist on several servers;
- Uses datetime as a global version indicator;
- Is based on the primitives of the Web: resource, resource state, representation, content negotiation, link.
Since Memento’s access approach is distributed, and is based on Web primitives, it scales like the Web.
original resource and versions
bridge from present to past
bridge from past to present
memento
multiple archives
original resource gone
original resource’s server gone
original resource provides no link
Memento has impact

on many levels
Improved access to digital memory

Existing Web archives become seamlessly accessible from the browser: original URI + time.

Versions become uniformly accessible irrespective of the nature of the hosting system: original URI + time.

This has significant impact for the Web of Documents, and is game changing for the Web of Data.
This capability will change user expectations regarding access to versions.

The changed user expectation will yield increased awareness regarding the need to archive versions.

The result will be a Web with a well-managed memory instead of an amnesiac one.
Memento directly helps to make our digital memory accessible.

Memento indirectly helps to expand our digital memory.

http://mementoweb.org/