

Memento

Accessing Resource Versions Using Datetime Negotiation



Herbert Van de Sompel
Robert Sanderson
Michael L. Nelson

<http://mementoweb.org/>

Memento is funded by
The Library of Congress

<http://www.slideshare.net/hvdsomp/memento-updated-technical-details-may-2011>



Memento
SWIB 2011, Hamburg, Germany, November 2011



Memento wants to make it Easy to navigate the Web of the Past

Technical Specification

<https://datatracker.ietf.org/doc/draft-vandesompe-memento/>



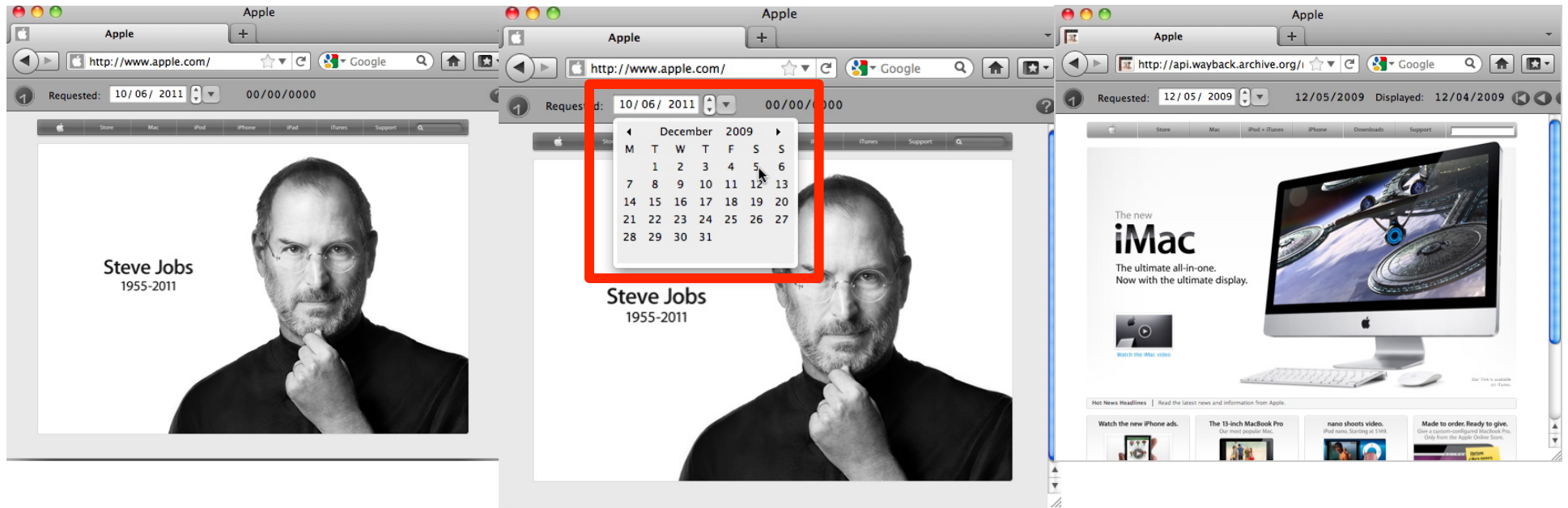
Memento
SWIB 2011, Hamburg, Germany, November 2011²



Apple
Today

Select Date
Dec 5 2009

Apple
Dec 5 2009



From
Internet Archive



Memento
SWIB 2011, Hamburg, Germany, November 2011

Memento achieves this by introducing
a uniform version access capability to
integrate the past and current Web



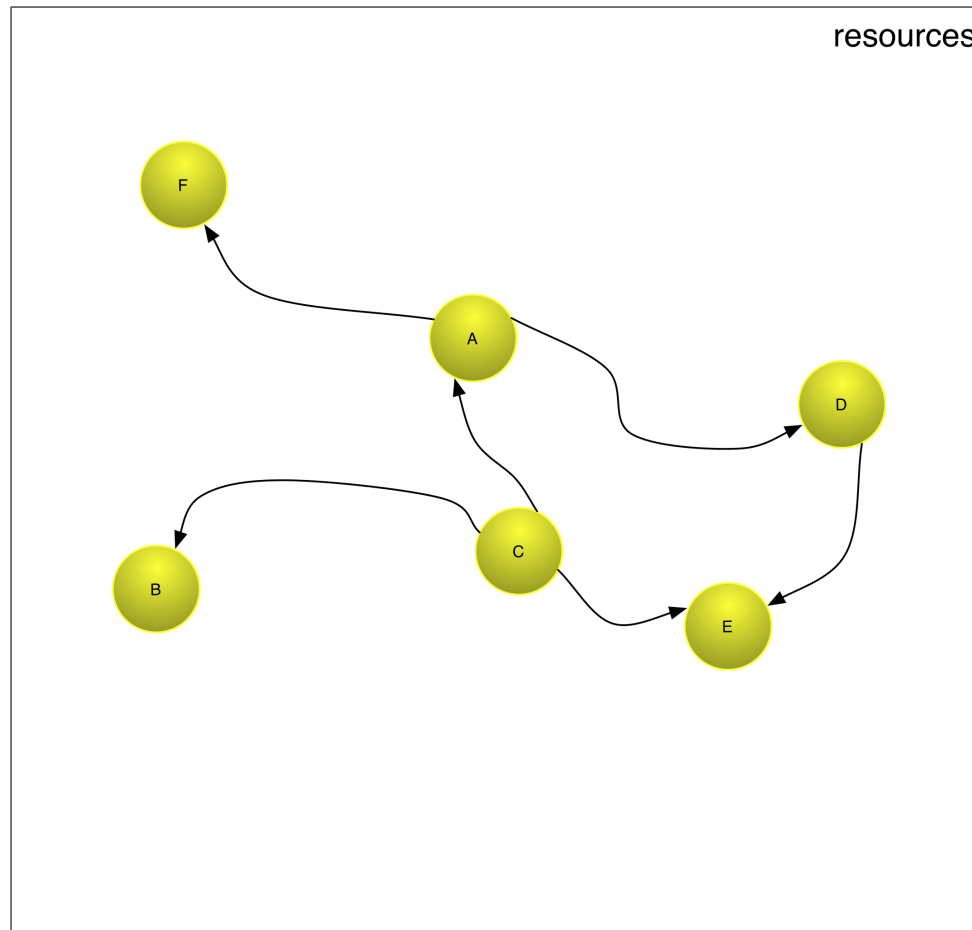
Problem Statement ...



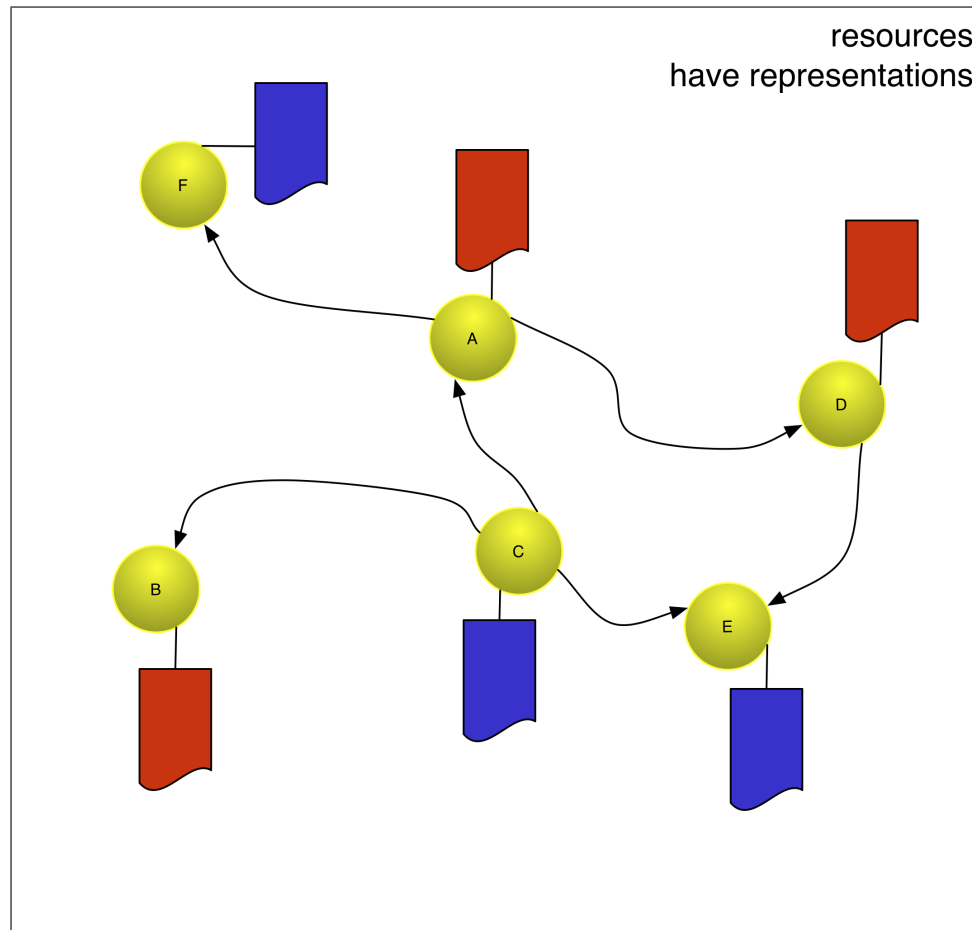
Memento
SWIB 2011, Hamburg, Germany, November 2011 ⁵



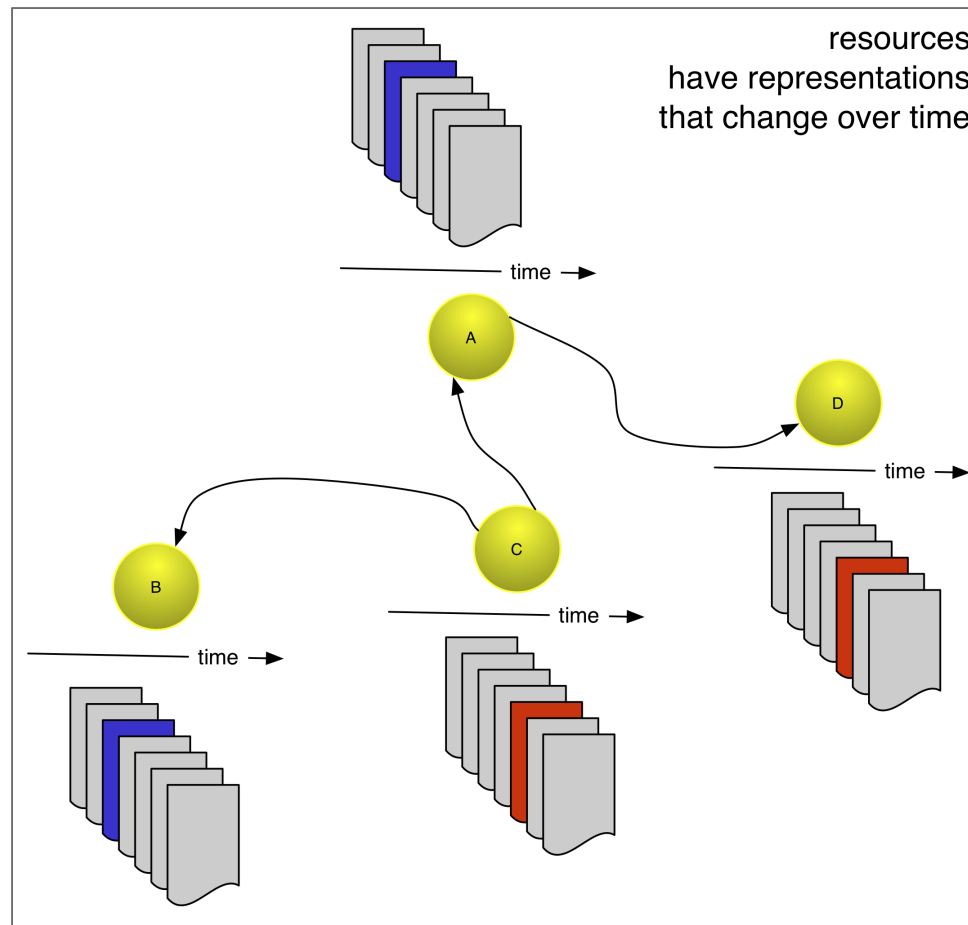
Resources



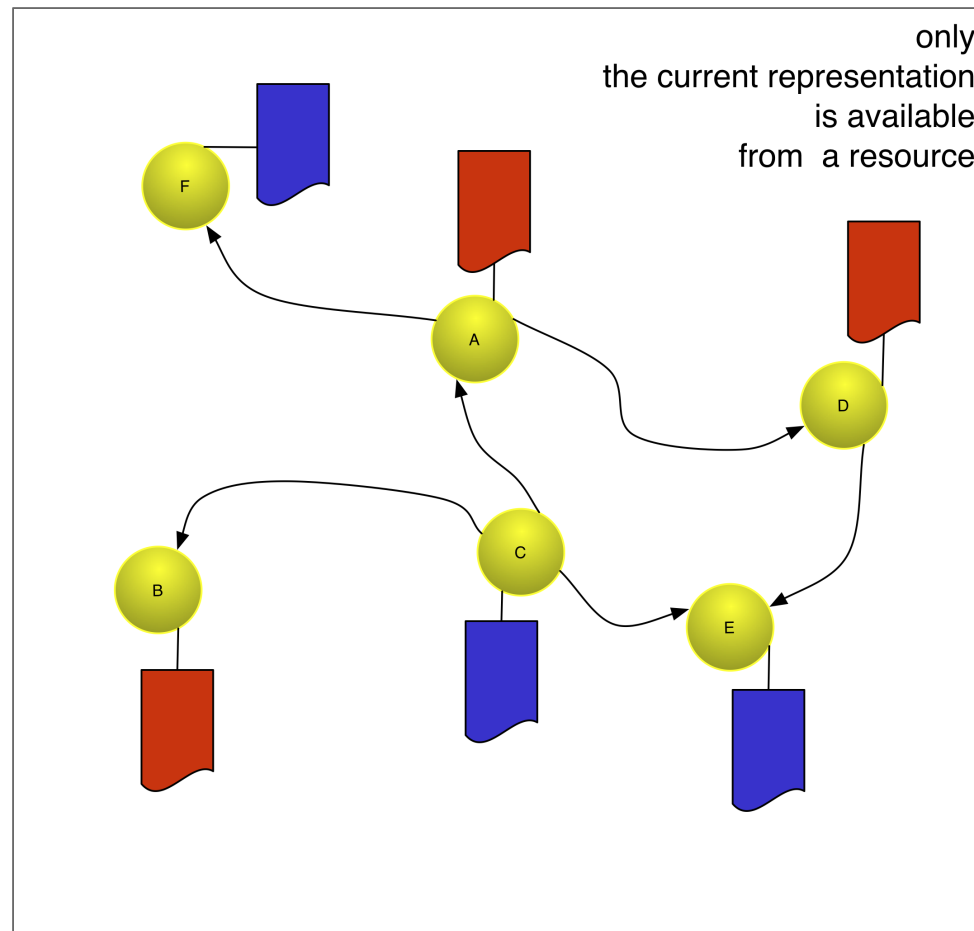
Resources have Representations



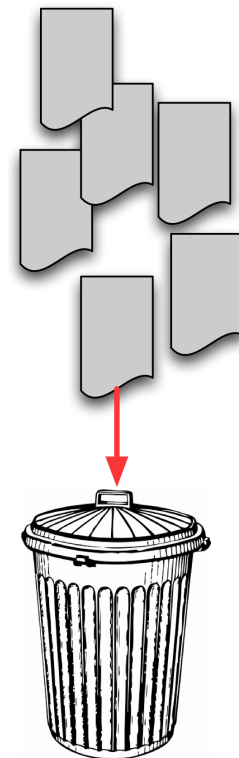
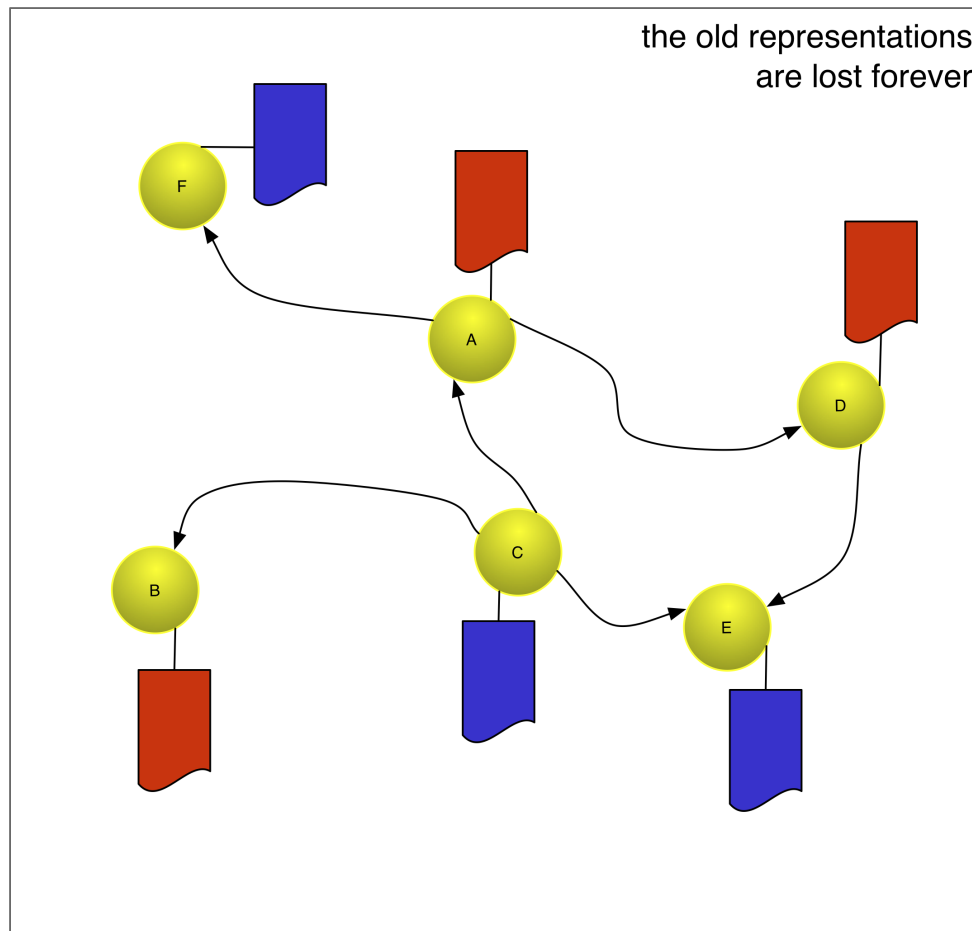
Resources have Representations that Change over Time



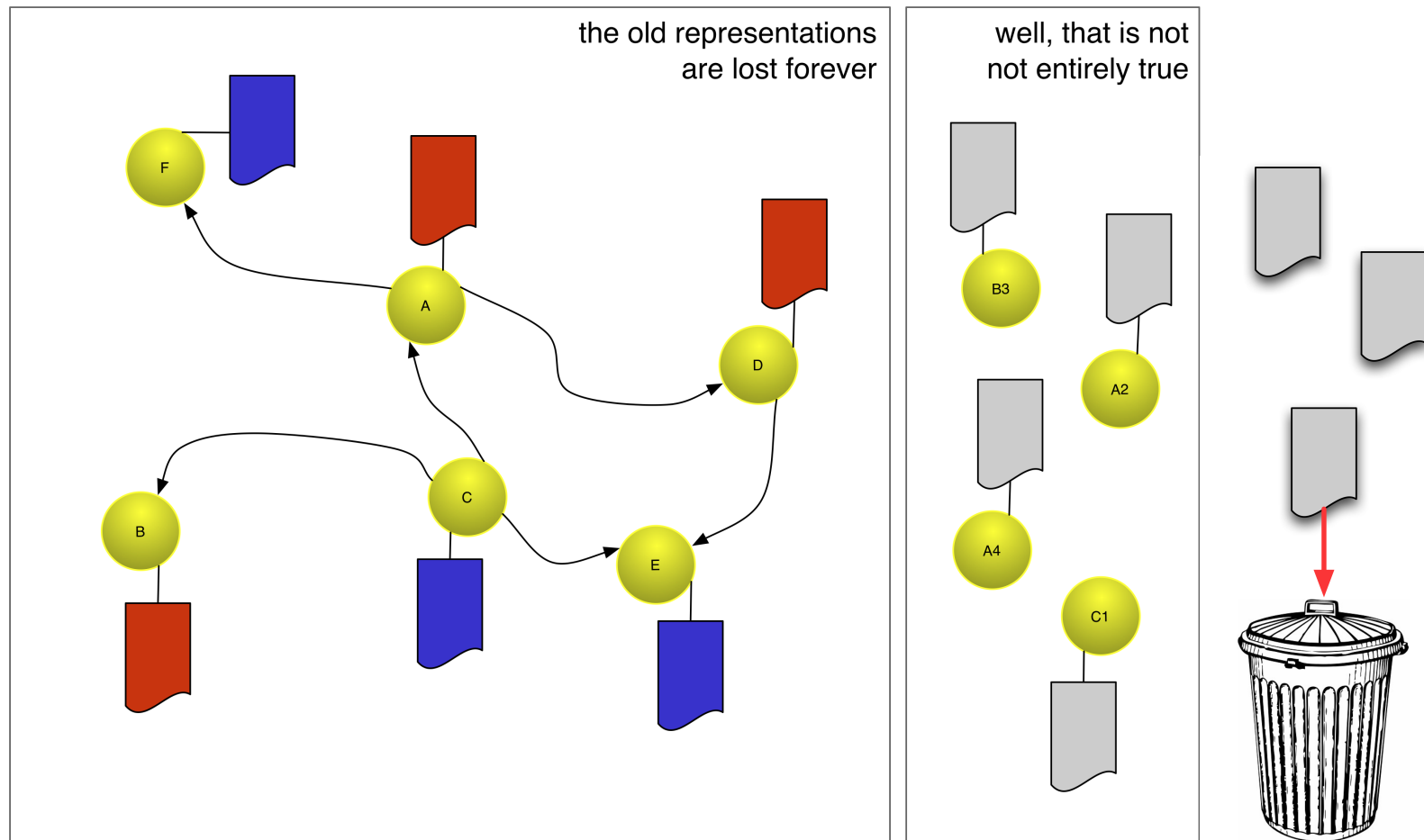
Only the Current Representation is Available from a Resource



Old Representations are Lost Forever



Archived/Version Resources Exist



Resource Versions on the Web



- Content Management Systems
- Web Archives
- Transactional archives
- Search engine caches
- ...



Sep 11 2001, 20:36:10 UTC

Dec 20 2001, 4:51:00 UTC

Resource Versions

The screenshot shows the CNN.com website's 'SPECIAL REPORT' section for the September 11 attacks. The main headline is 'AMERICA UNDER ATTACK' with a sub-headline 'CNN EXCLUSIVE'. A large image shows the World Trade Center towers on fire. The text describes the attacks at 8:45 a.m. EDT. The page includes a 'COMPLETE COVERAGE' section with links to various news stories, a 'U.S. SCENE' section, a 'WORLD SCENE' section, and a 'RESOURCES AND RESCUE' section. There are also sections for 'VIDEO', 'PHOTO GALLERY', 'CHRONOLOGY', and 'EXTRA INFO'.

The screenshot shows the Wikipedia article 'September 11 attacks' as it appeared on December 20, 2001. The article is marked as an 'old revision' and includes a notice that it differs significantly from the current revision. The article text describes the attacks on the World Trade Center and the Pentagon, the casualties, and the global impact. It also mentions the U.S. government's response and the subsequent attack on Afghanistan. The page includes a navigation sidebar with links to various Wikipedia features and a search bar.

<http://web.archive.org/web/20010911203610/http://www.cnn.com/> archived resource for <http://cnn.com>

http://en.wikipedia.org/w/index.php?title=September_11_attacks&oldid=282333 archived resource for http://en.wikipedia.org/wiki/September_11_attacks



Memento
SWIB 2011, Hamburg, Germany, November 2011 13



Versions are Not Integrated with the Web



Memento
SWIB 2011, Hamburg, Germany, November 2011 14



Architecture of the Web and Resource Versions



Architecture of the World Wide Web, Volume One

W3C Recommendation 15 December 2004

This version:

<http://www.w3.org/TR/2004/REC-webarch-20041215/>

Latest version:

<http://www.w3.org/TR/webarch/>

Previous version:

<http://www.w3.org/TR/2004/PR-webarch-20041105/>

Editors:

[Ian Jacobs](#), W3C

Norman Walsh, Sun Microsystems, Inc.

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.



Memento

SWIB 2011, Hamburg, Germany, November 2011

15



Finding Resource Versions

able here 6 months after collection. [See FAQ.](#)

Search Results for Jan 01, 1996 - May 14, 2009							
2000	2001	2002	2003	2004	2005	2006	2007
30 pages	2734 pages	231 pages	156 pages	625 pages	1186 pages	886 pages	953 pages
Jun 20, 2000	Mar 01, 2001	Jan 23, 2002	Jan 29, 2003	Jan 01, 2004	Jan 01, 2005	Jan 01, 2006	Jan 01, 2007
Jun 21, 2000	Mar 02, 2001	Feb 02, 2002	Jan 31, 2003	Jan 02, 2004	Jan 01, 2005	Jan 01, 2006	Jan 01, 2007
Jun 21, 2000	Mar 05, 2001	Feb 09, 2002	Feb 01, 2003	Jan 03, 2004	Jan 02, 2005	Jan 02, 2006	Jan 02, 2007
Jun 21, 2000	Mar 07, 2001	Feb 10, 2002	Feb 01, 2003	Jan 10, 2004	Jan 02, 2005	Jan 02, 2006	Jan 02, 2007
Jun 21, 2000	Mar 09, 2001	Feb 11, 2002	Feb 02, 2003	Jan 11, 2004	Jan 03, 2005	Jan 03, 2006	Jan 02, 2007
Jul 06, 2000	Mar 31, 2001	Feb 13, 2002	Feb 02, 2003	Jan 12, 2004	Jan 04, 2005	Jan 03, 2006	Jan 03, 2007
Jul 07, 2000	Apr 10, 2001	Feb 13, 2002	Feb 03, 2003	Jan 12, 2004	Jan 05, 2005	Jan 03, 2006	Jan 03, 2007
Jul 11, 2000	Apr 13, 2001	Feb 14, 2002	Feb 05, 2003	Jan 13, 2004	Jan 06, 2005	Jan 04, 2006	Jan 04, 2007
Aug 04, 2000	Apr 13, 2001	Feb 15, 2002	Feb 06, 2003	Jan 15, 2004	Jan 06, 2005	Jan 04, 2006	Jan 05, 2007
Aug 15, 2000	Apr 13, 2001	Feb 16, 2002	Feb 07, 2003	Jan 16, 2004	Jan 06, 2005	Jan 04, 2006	Jan 05, 2007
Aug 15, 2000	Apr 14, 2001	Feb 22, 2002	Feb 10, 2003	Jan 20, 2004	Jan 07, 2005	Jan 05, 2006	Jan 06, 2007
Aug 15, 2000	Apr 14, 2001	Feb 22, 2002	Feb 11, 2003	Jan 20, 2004	Jan 08, 2005	Jan 05, 2006	Jan 06, 2007
Aug 16, 2000	Apr 18, 2001	Feb 23, 2002	Feb 12, 2003	Jan 20, 2004	Jan 09, 2005	Jan 05, 2006	Jan 07, 2007
Aug 16, 2000	May 03, 2001	Feb 23, 2002	Feb 12, 2003	Jan 24, 2004	Jan 09, 2005	Jan 06, 2006	Jan 07, 2007
Aug 17, 2000	May 03, 2001	Feb 23, 2002	Jan 27, 2004	Jan 27, 2004	Jan 10, 2005	Jan 06, 2006	Jan 07, 2007
Aug 17, 2000	May 03, 2001	Feb 24, 2002	Feb 15, 2003	Jan 28, 2004	Jan 10, 2005	Jan 06, 2006	Jan 07, 2007
Aug 18, 2000	May 03, 2001	Feb 26, 2002	Mar 20, 2003	Jan 30, 2004	Jan 10, 2005	Jan 06, 2006	Jan 07, 2007
Aug 28, 2000	May 03, 2001	Feb 27, 2002	Mar 21, 2003	Jan 30, 2004	Jan 11, 2005	Jan 07, 2006	Jan 08, 2007
Sep 20, 2000	May 03, 2001	Mar 27, 2002	Mar 21, 2003	Jan 31, 2004	Jan 11, 2005	Jan 07, 2006	Jan 08, 2007
Sep 30, 2000	May 03, 2001	Mar 29, 2002	Mar 24, 2003	Feb 02, 2004	Jan 12, 2005	Jan 07, 2006	Jan 08, 2007
Oct 18, 2000	May 04, 2001	May 25, 2002	Mar 25, 2003	Feb 02, 2004	Jan 12, 2005	Jan 07, 2006	Jan 08, 2007
Nov 09, 2000	May 04, 2001	May 27, 2002	Mar 26, 2003	Feb 03, 2004	Jan 13, 2005	Jan 08, 2006	Jan 08, 2007
Nov 10, 2000	May 04, 2001	May 27, 2002	Mar 29, 2003	Feb 04, 2004	Jan 13, 2005	Jan 09, 2006	Jan 08, 2007
Nov 10, 2000	May 04, 2001	May 28, 2002	Mar 29, 2003	Feb 07, 2004	Jan 13, 2005	Jan 09, 2006	Jan 08, 2007
Nov 17, 2000	May 04, 2001	May 29, 2002	Mar 30, 2003	Feb 10, 2004	Jan 14, 2005	Jan 10, 2006	Jan 08, 2007
Nov 18, 2000	May 04, 2001	May 30, 2002	Mar 31, 2003	Feb 10, 2004	Jan 14, 2005	Jan 10, 2006	Jan 09, 2007
Dec 04, 2000	May 04, 2001	May 30, 2002	Apr 01, 2003	Feb 11, 2004	Jan 14, 2005	Jan 10, 2006	Jan 09, 2007
Dec 04, 2000	May 04, 2001	May 30, 2002	Apr 04, 2003	Feb 13, 2004	Jan 15, 2005	Jan 10, 2006	Jan 09, 2007
Dec 04, 2000	May 05, 2001	May 31, 2002	Apr 04, 2003	Feb 13, 2004	Jan 15, 2005	Jan 10, 2006	Jan 09, 2007
Dec 04, 2000	May 05, 2001	Jun 02, 2002	Apr 05, 2003	Feb 14, 2004	Jan 15, 2005	Jan 10, 2006	Jan 10, 2007
	May 05, 2001	Jun 02, 2002	Apr 08, 2003	Feb 14, 2004	Jan 16, 2005	Jan 10, 2006	Jan 10, 2007
	May 05, 2001	Jun 03, 2002	Apr 09, 2003	Feb 17, 2004	Jan 16, 2005	Jan 11, 2006	Jan 11, 2007
	May 05, 2001	Jun 03, 2002	Apr 09, 2003	Feb 18, 2004	Jan 17, 2005	Jan 11, 2006	Jan 11, 2007
	May 05, 2001	Jun 04, 2002	Apr 09, 2003	Feb 20, 2004	Jan 17, 2005	Jan 11, 2006	Jan 12, 2007
	May 05, 2001	Jun 06, 2002	Apr 11, 2003	Feb 21, 2004	Jan 18, 2005	Jan 11, 2006	Jan 12, 2007
	May 05, 2001	Jul 01, 2002	Apr 20, 2003	Feb 28, 2004	Jan 18, 2005	Jan 11, 2006	Jan 12, 2007
	May 05, 2001	Jul 02, 2002	Apr 23, 2003	Mar 02, 2004	Jan 19, 2005	Jan 11, 2006	Jan 12, 2007
	May 05, 2001	Jul 03, 2002	Apr 23, 2003	Mar 07, 2004	Jan 19, 2005	Jan 12, 2006	Jan 13, 2007

Go to <http://www.archive.org/> and search <http://cnn.com>

On http://web.archive.org/web/*/http://cnn.com, select desired datetime



Finding Resource Versions



The screenshot shows the Wikipedia article for "September 11 attacks". The article text describes the coordinated suicide attacks by Al-Qaeda on September 11, 2001, in New York City and Arlington, Virginia. It mentions the hijacking of four commercial passenger jet airliners, the collapse of the Twin Towers of the World Trade Center, and the crash of a third airliner into the Pentagon. The article also notes the deaths of 2,993 people, including the hijackers, and the overwhelming majority of casualties were civilians. It mentions the "War on Terrorism" launched by the United States in response to the attacks.



The screenshot shows the "Revision history of September 11 attacks" page on Wikipedia. It displays a list of revisions, including the current version (cur) and previous versions (prev). The list includes the date, time, and user for each revision, along with the size of the edit. The current version is 18:20, 9 November 2009 by InnerParty (talk | contribs) (133,903 bytes). The previous version is 23:01, 8 November 2009 by InnerParty (talk | contribs) (133,650 bytes). The list also includes a "Compare selected revisions" button and a "Browse history" button.

Go to

http://en.wikipedia.org/wiki/September_11_attacks
and click History

Browse History



Memento

SWIB 2011, Hamburg, Germany, November 2011

17



Dec 20 2001, 4:51:00 UTC

current

Navigating Resource Versions

September 11 attacks

From Wikipedia, the free encyclopedia

This is an old revision of this page, as edited by The Cunctator (talk | contris) at 04:51, 20 December 2001. It may differ significantly from the current revision.

(diff) ← Previous revision | Current revision (diff) | Newer revision → (diff)

In Memoriam, September 11, 2001

On the morning of **September 11, 2001**, what might well be the most devastating **terrorist attack** in the history of the world occurred concurrently in **New York City**, **W** jets were **hijacked** and then deliberately crashed into th of the World Trade Center subsequently collapsed, and fire. **Casualties** are expected to be in the thousands: 265 on the planes; about 3000 people (early estimates ranged as high as 6500 people), including hundreds of firefighters who had rushed in, at the World Trade Center; and 125 at the Pentagon.

Some passengers on the doomed flights were able to make phone calls reporting on events on board. They reported that there were more than one hijacker on each plane and that they took control of the planes using **box-cutter knives**. It appears that the passengers on the fourth jet tried to overpower the hijackers and that the plane crashed in a sparsely populated area as a result, thereby missing its intended target, which may have been the **White House**.

The attack had immediate and deep global **political effects** and **economic effects** and inspired an international outpouring of **memorials and services**. The massive undertaking of **rescue and recovery**, and of providing **assistance** to the **survivors** and victims, is ongoing. There will be a great need for **donations** for a long time.

Though no group has claimed **responsibility**, the U.S. government immediately launched a full-scale **response**, stating its intentions to go to war against those responsible. On October 7, a coalition led by the United States launched an attack in Afghanistan. See **2001 U.S. Attack on Afghanistan**.

Following the attack, the United States has been on heightened alert for new terrorist attacks. In late September, **cases of anthrax** started breaking out; although, no connection to the September 11 attack has been found. See **2001 anthrax attack**.

The Pentagon

From Wikipedia, the free encyclopedia

Coordinates: 38°52′15.56″N 77°3′21.46″W﻿ / ﻿

*This article is about the United States military building. For the shape, see **Pentagon**.*

The Pentagon is the headquarters of the **United States Department of Defense**, located in **Arlington, Virginia**. As a symbol of the U.S. military, "the Pentagon" is often used **metonymically** to refer to the Department of Defense rather than the building itself.

Designed by the American architect **George Bergstrom** (1876 – 1955), and built by **Philadelphia, Pennsylvania**, general contractor **John McShain**, the building was dedicated on January 15, 1943, after ground was broken for construction on September 11, 1941. General **Brehon Somervell** provided the major motive power behind the project;^[1] Colonel **Leslie Groves** and Major **Clarence Renshaw** were responsible for overseeing the project for the Army.

The Pentagon is the world's largest office building by floor area^{[2][3]}, by a margin of about 35,580 square metres (383,000 sq ft). Approximately 23,000 military and civilian employees^[3] and about 3,000 non-defense support personnel work in the Pentagon. It has five sides, five floors above ground (plus two basement levels), and five ring corridors per floor with a total of 17.5 miles (28.2 km)^[3] of corridors. The Pentagon includes a five-acre (20,000 m²) central plaza, which is shaped like a pentagon and informally known as "**ground zero**", a nickname originating during the **Cold War** and based on the presumption that the **Soviet Union** would target one or more nuclear missiles at this central location.

On **September 11, 2001**, hijacked **American Airlines Flight 77** was crashed into the western side of the

The Pentagon
Arlington, Virginia

Southwest view of the **Pentagon** with the **Potomac River** and **Washington Monument** in background.

Shown on District of Columbia map

Type	Headquarters
Coordinates	38°52′15.56″N 77°3′21.46″W﻿ / ﻿
Built	1943 (broke ground on September 11, 1941)
Height	Five floors above ground and Two floors below ground
In use	1943–present

http://en.wikipedia.org/w/index.php?title=September_11_attacks&oldid=282333 archived resource for http://en.wikipedia.org/wiki/September_11_attacks3

http://en.wikipedia.org/wiki/The_Pentagon



Memento

SWIB 2011, Hamburg, Germany, November 2011

18



Sep 11 2001, 20:36:10 UTC

Sep 11 2001, 21:38:55 UTC

Navigating Resource Versions

CNN.com SPECIAL REPORT

Updated: 5:42 p.m. EDT -- 11 September 2001

AMERICA UNDER ATTACK

CNN EXCLUSIVE

COMPLETE COVERAGE

- World Trade Center towers collapse after hit
- Crash destroys part of Pentagon
- White House, Capitol evacuated
- American, United confirm losing planes
- Bush: 'We'll hunt them down' | Statement
- U.S. officials: More attacks can't be ruled out
- Attack strikes financial markets

U.S. SCENE

- FAA grounds all U.S. flights until noon Wednesday
- U.S. military on 'high alert'
- Major sports events postponed

WORLD SCENE

- Taliban statement: 'Afghanistan feels your pain'
- World in shock | Evacuations amid chaos
- Middle East, Asia condemn attacks

RESOURCES AND RESCUE

Giuliani: 1,500 'walking wounded' in New York
[Emergency contact information](#)

Terrorists strike New York, D.C.

At 8:45 a.m. EDT, the first of two airliners crashed into the World Trade Center, opening a horrifying and apparently coordinated terrorist attack on the United States, which saw the collapse of the two 110-story towers into surrounding Manhattan streets and a later attack on the Pentagon.

[FULL STORY >>](#)

[SPECIAL REPORT >>](#)

VIDEO

Watch video of second plane crash at World Trade Center
[More audio/video >>](#)

PHOTO GALLERY

Aftermath of the attacks in New York and Washington

CHRONOLOGY

At 8:45 a.m. EDT, day of disaster begins
[Map of attacks](#)

EXTRA INFO

[World Trade Center, Pentagon, 767](#)

CNN.com / SPACE

Updated: 04:56 PM EDT 20:56 GMT -- 11 Sep 2001

[Cool star chills stellar theories](#)

TOP STORIES

- NASA shuts down in wake of attacks
- Sleuths try to unmask new spy satellites
- Probe limps to close encounter with comet
- Report: Plane tracking satellites can be jammed

[More News >](#)

Cool star chills stellar theories

A star besieged by unbelievably hot waves of nuclear explosions is ten times cooler than expected, forcing astronomers to rethink their ideas about stellar physics.

[FULL STORY >](#)

ALPHA EXTRA

Space correspondent Miles O'Brien on the space station tour

SCENES FROM SPACE

A NASA telescope is marking a milestone

RESOURCES:

SPACE AGENCY NEWS

The latest news releases from worldwide space agencies.

[Hit the links](#)

INTERACTIVE LIBRARY

Your window to the universe -- launch video, 360-degree photo tours and more.

NASA WEBCAMS

Live pictures of NASA projects.

[Launch it](#)

SPACE CALENDARS

Calendars of upcoming launches and other events from worldwide space agencies.

[Calendar list](#)

SPACE PROBES

A guide to active unmanned missions.

[Probe list](#)

MORE NEWS

Space program subject of 'Antarctica' pilot

<http://web.archive.org/web/20010911203610/http://www.cnn.com/> archived resource for <http://cnn.com>

<http://web.archive.org/web/20010911213855/www.cnn.com/TECH/space/>



Memento
SWIB 2011, Hamburg, Germany, November 2011

19



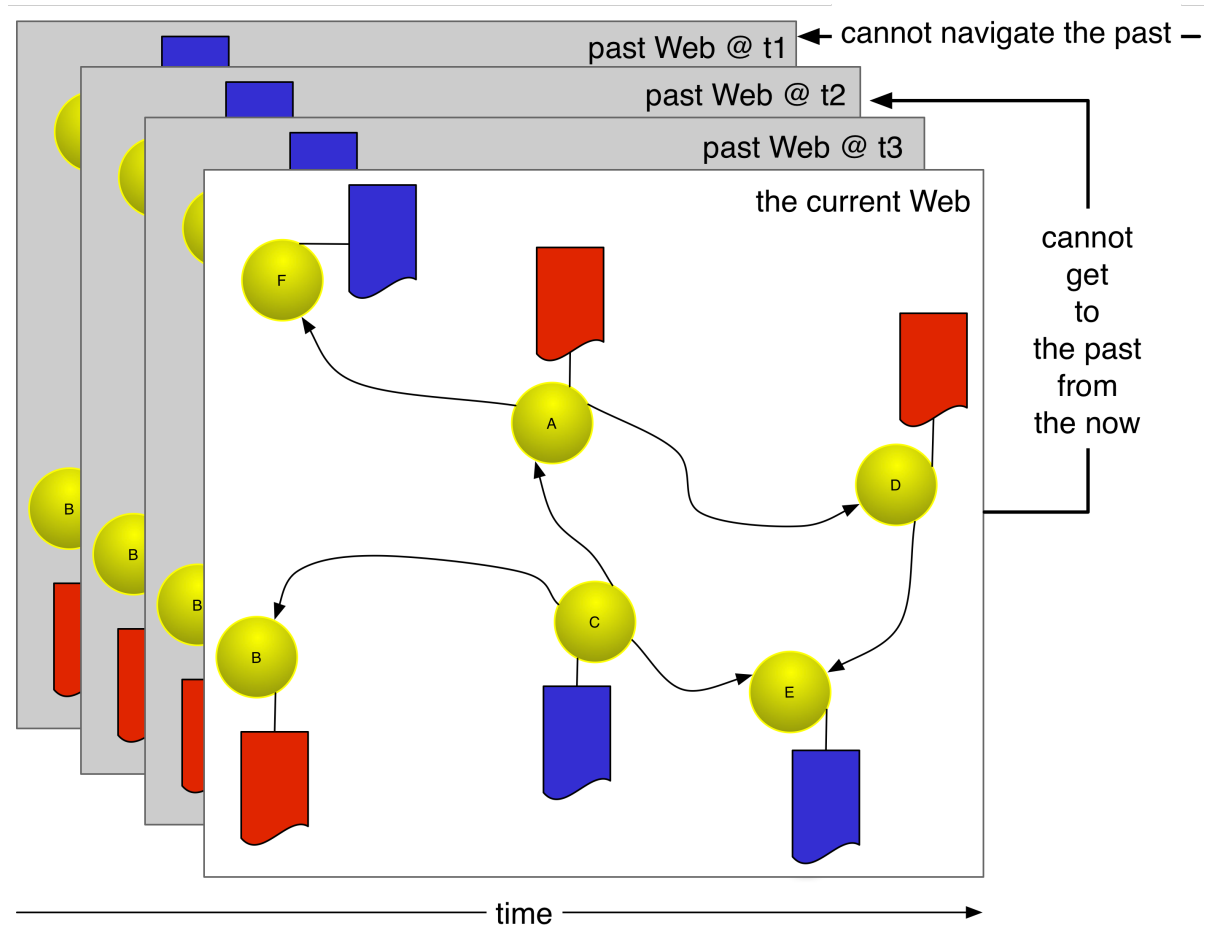
Versions are Not Integrated with the Web



- Functionally:
 - Need to know about archives
 - Need to search each of them individually
 - Cannot consistently navigate in the past within an archive
 - Cannot navigate across archives
- Architecturally:
 - 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



Memento Wants to Integrate the Past and Current Web



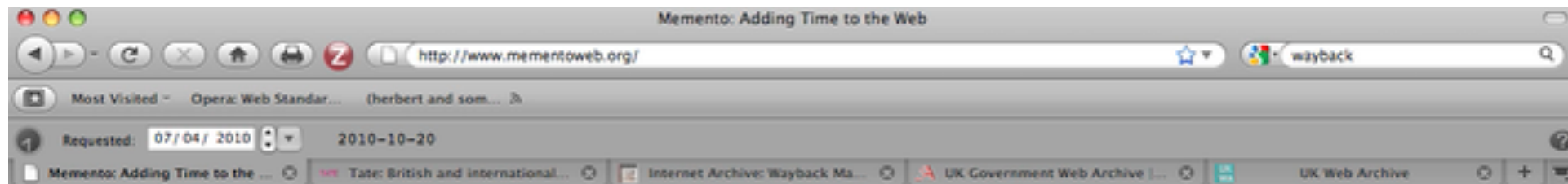
The Memento Framework:

Accessing Resource Versions with and without Memento



Memento
SWIB 2011, Hamburg, Germany, November 2011²²





Memento Adding Time to the Web

[About](#) [Demos](#) [Guide](#) [Tools](#)

Memento wants to make it as straightforward to access the Web of the past as it is to access the current Web.

If you know the URI of a Web resource, the technical framework proposed by Memento allows you to see a version of that resource as it existed at some date in the past, by entering that URI in your browser like you always do and by specifying the desired date in a browser plug-in. Or you can actually browse the Web of the past by selecting a date and clicking away. Whatever you land upon will be versions of Web resources as they were around the selected date. Obviously, this will only work if previous versions are available somewhere on the Web. But if they are, and if they are on servers that support the Memento framework, you will get to them.

At this point, there aren't any formal technical specifications detailing the Memento framework but we will get to that. For now, the information on this site should provide quite a good insight into how Memento is trying to change the Web by adding a time dimension to its most common protocol, HTTP. The [Guide to Memento](#) is a good entry point.

The Memento Framework: Protocol to Integrate Past and Current Web

Overview



Memento Framework



- Regards the Web as a big Content Management System
- Introduces a uniform capability to access versions on the Web
- Does not build new archives but leverages existing systems that host versions



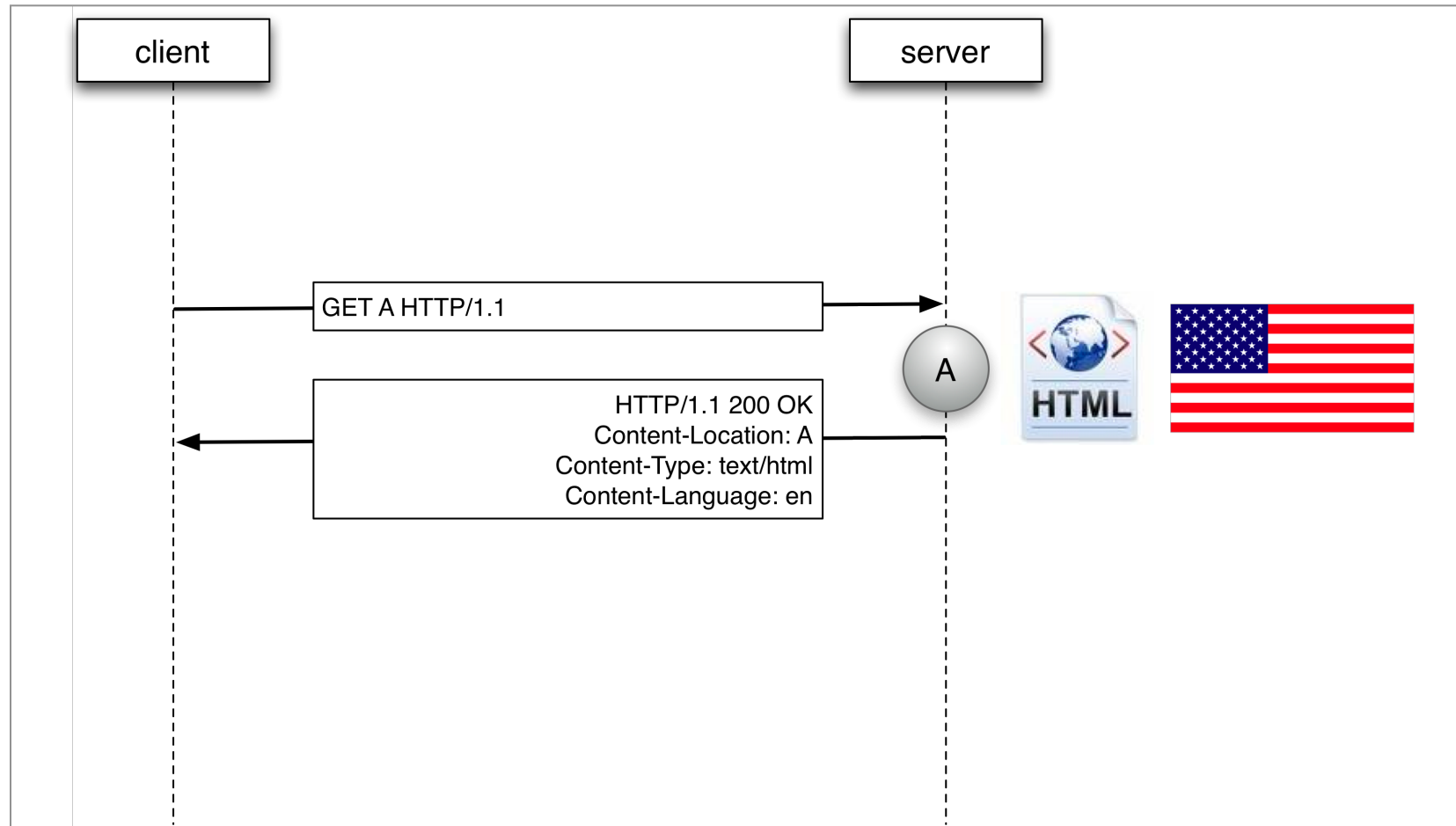
Memento Framework



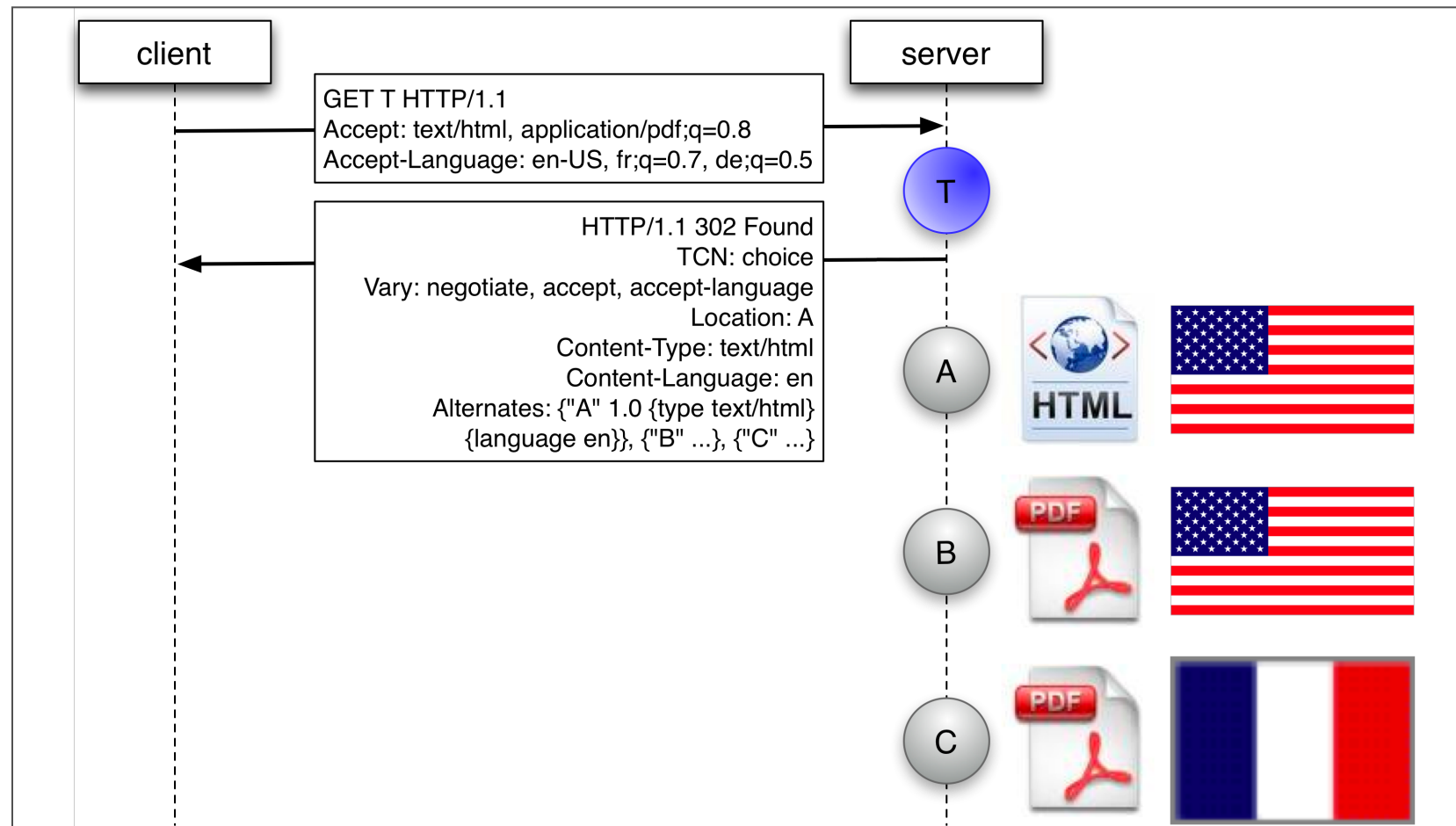
- Is distributed: versions may exist on several servers
- Uses time as a global version indicator
- Is based on the primitives of the Web: resource, resource state, representation, content negotiation, link



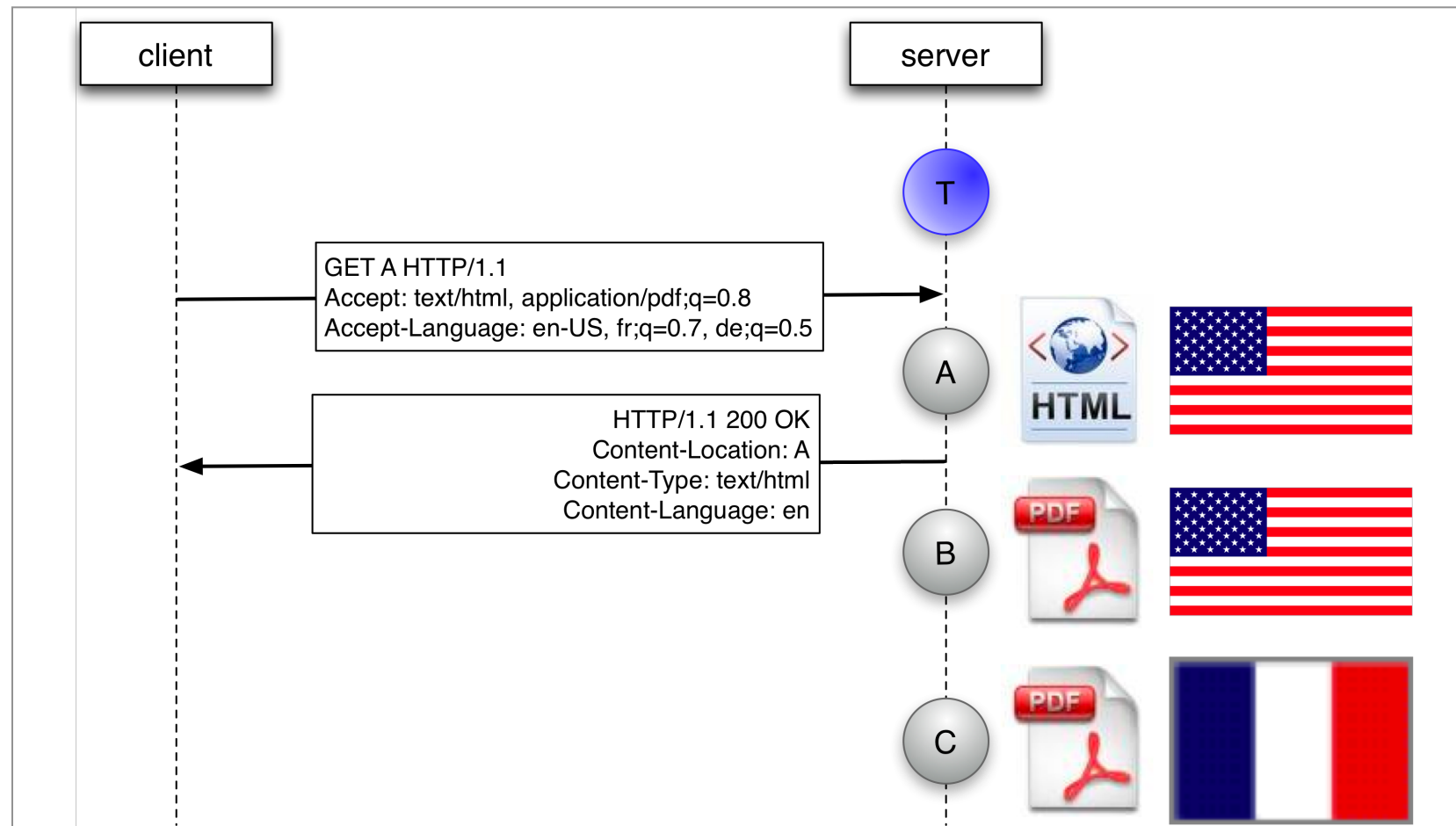
HTTP GET on URI-A



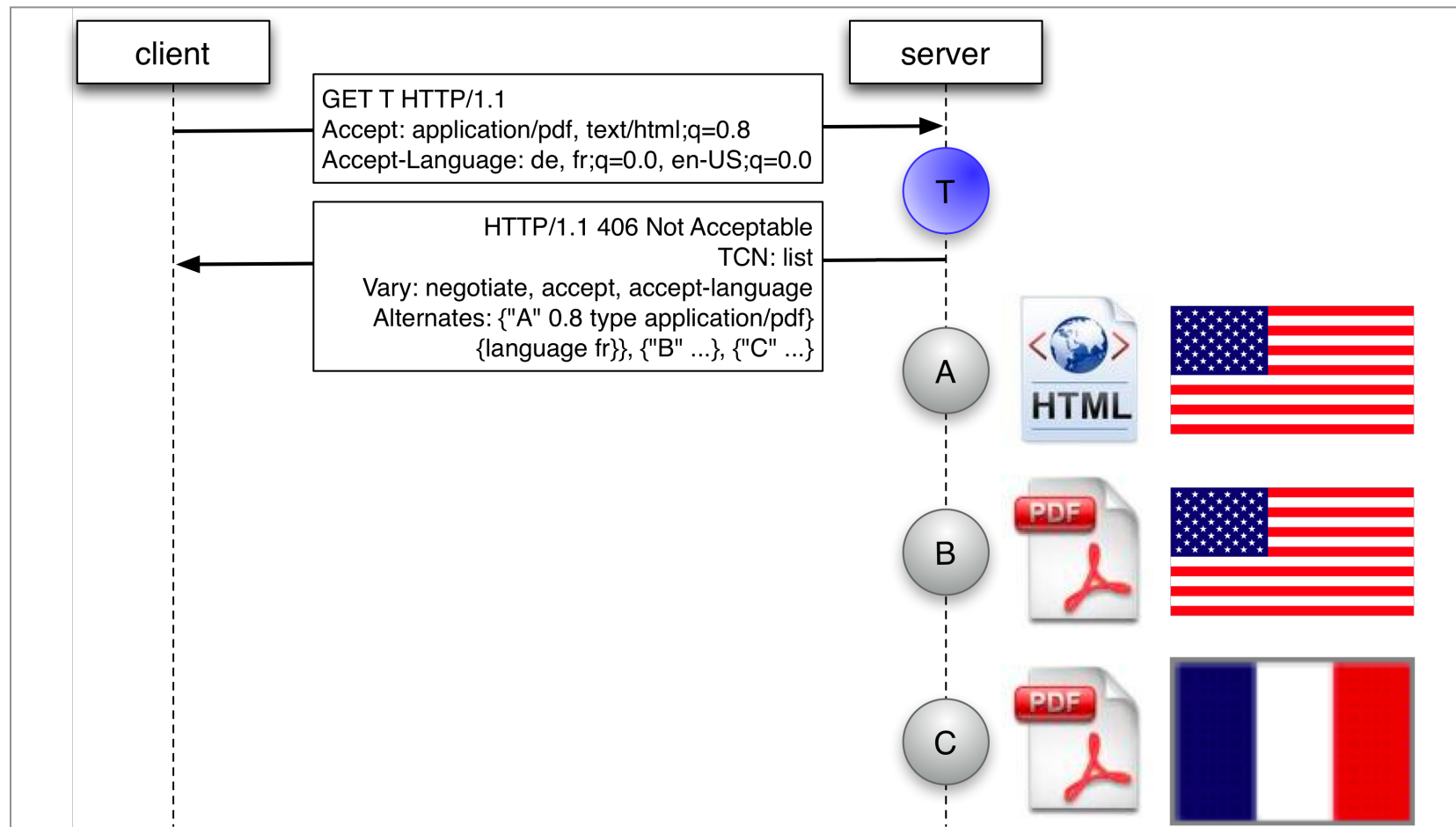
GET with conneg on URI-T – Server Choice – 302 Found – Step 1



GET with conneg on URI-T – Server Choice – 302 Found – Step 2



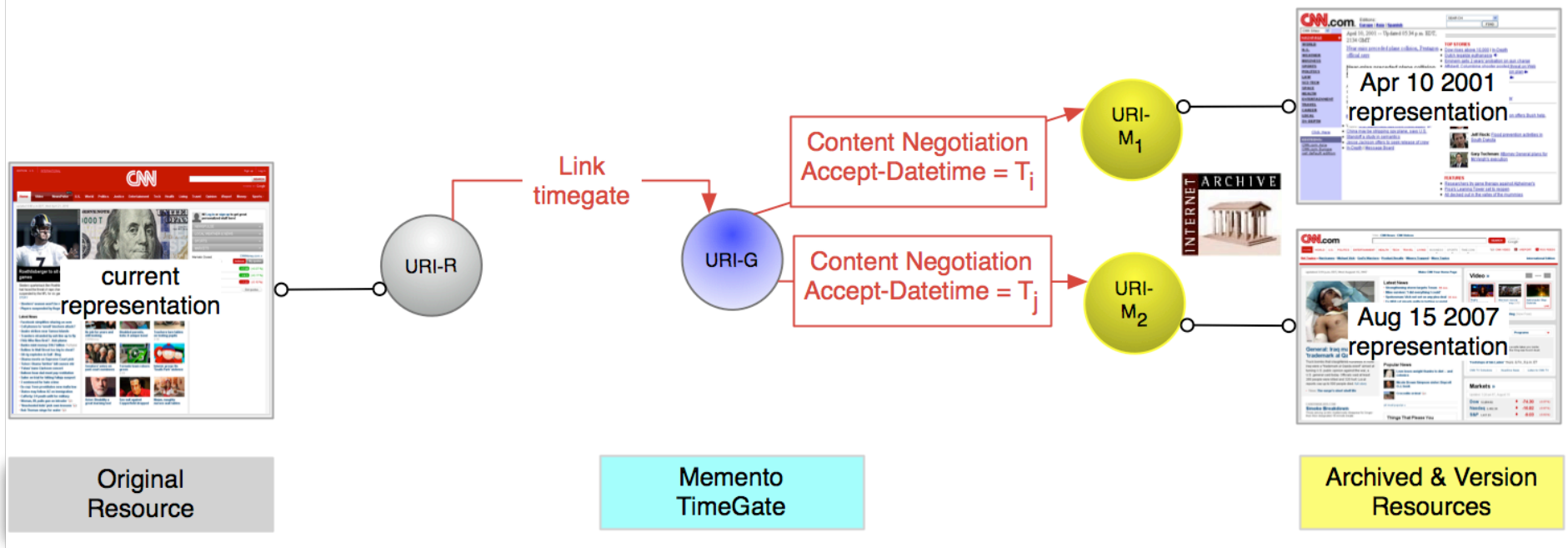
GET with conneg on URI-T – Server List – 406 Not Acceptable



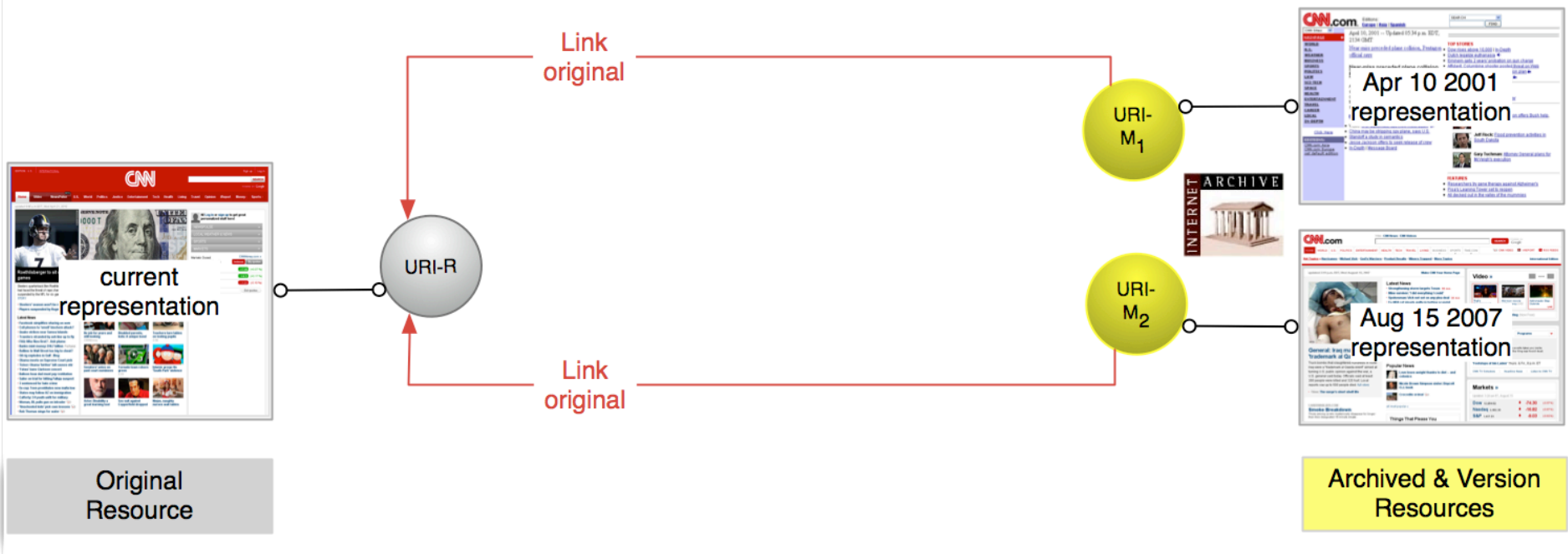
Original Resources and Mementos



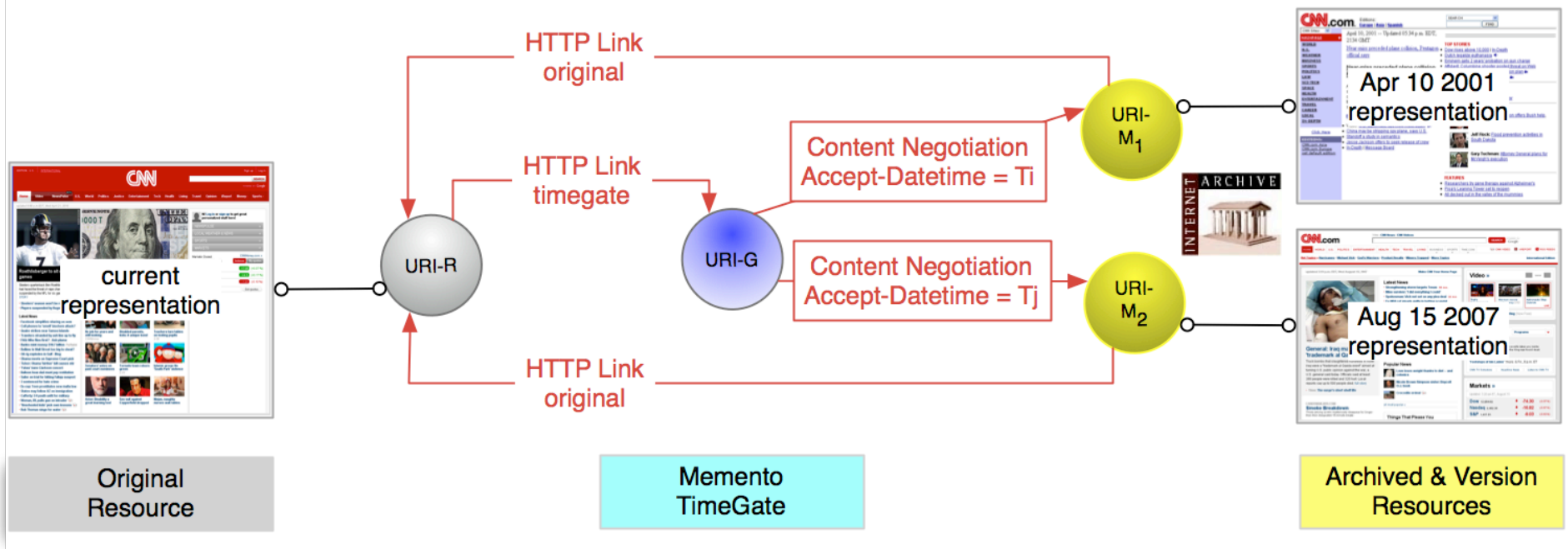
Bridge from Present to Past



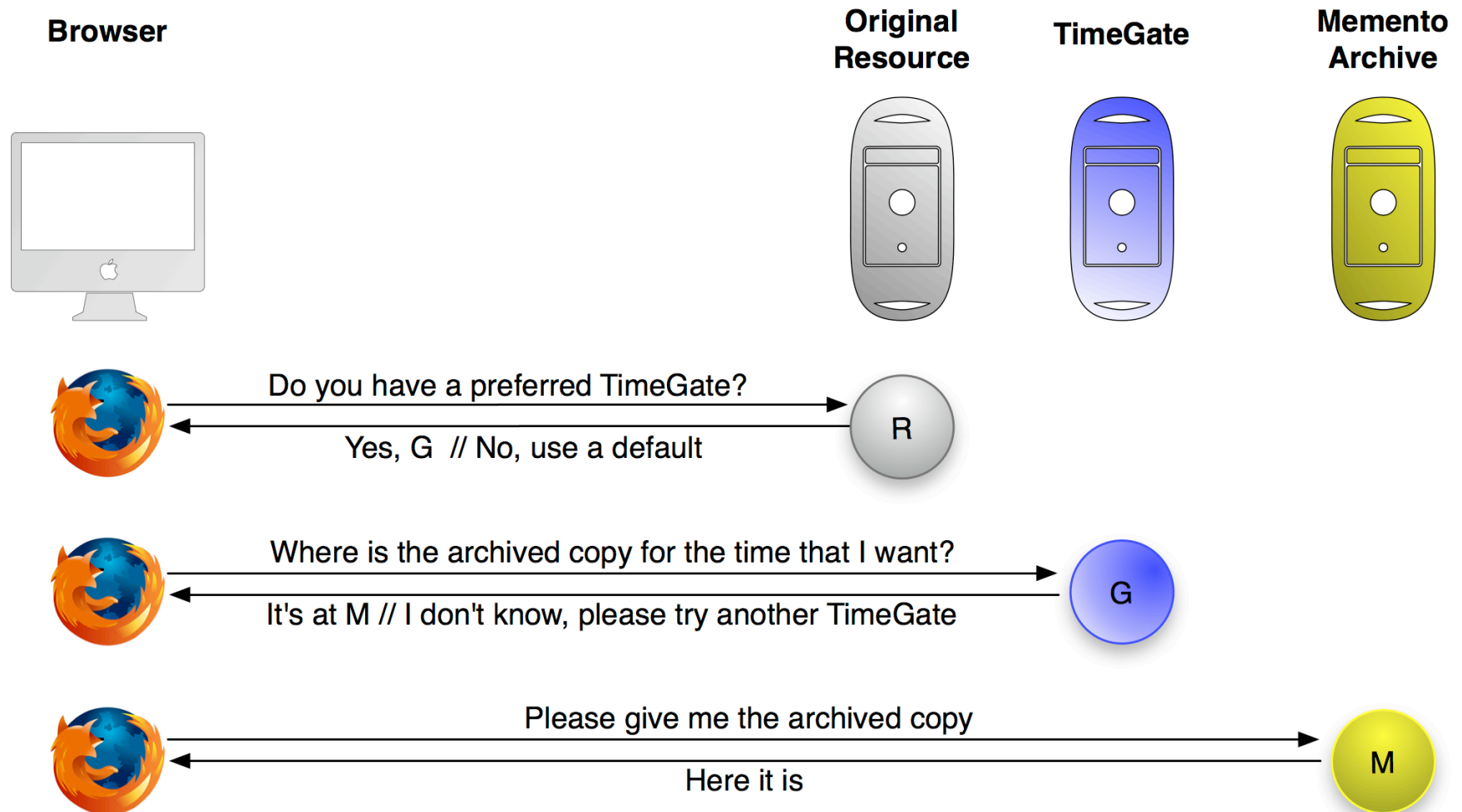
Bridge from Past to Present



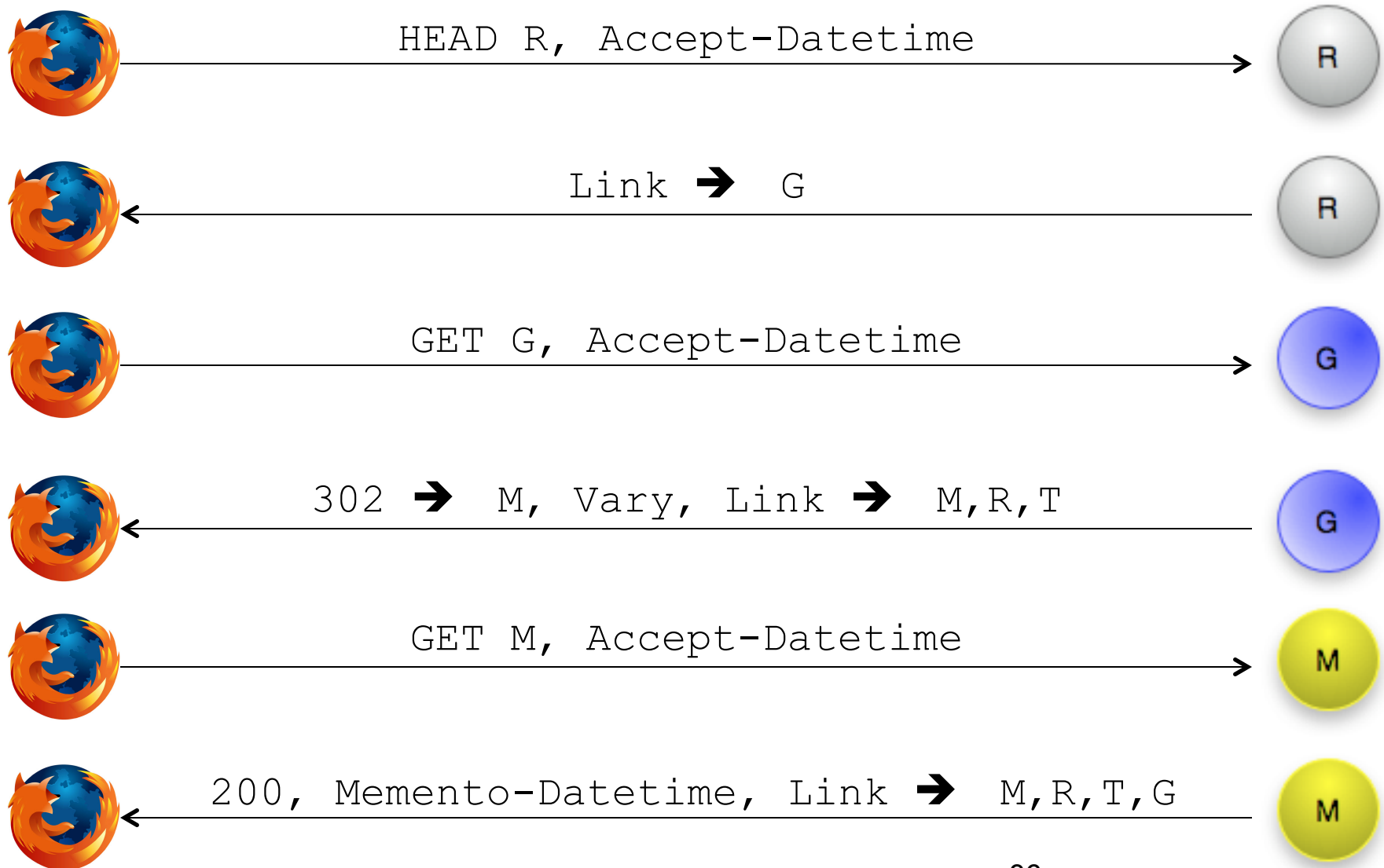
Memento Framework



Memento Client Server Interaction



Memento HTTP Flow



The Memento Framework:

Protocol to Integrate Past and Current Web

Interesting Cases

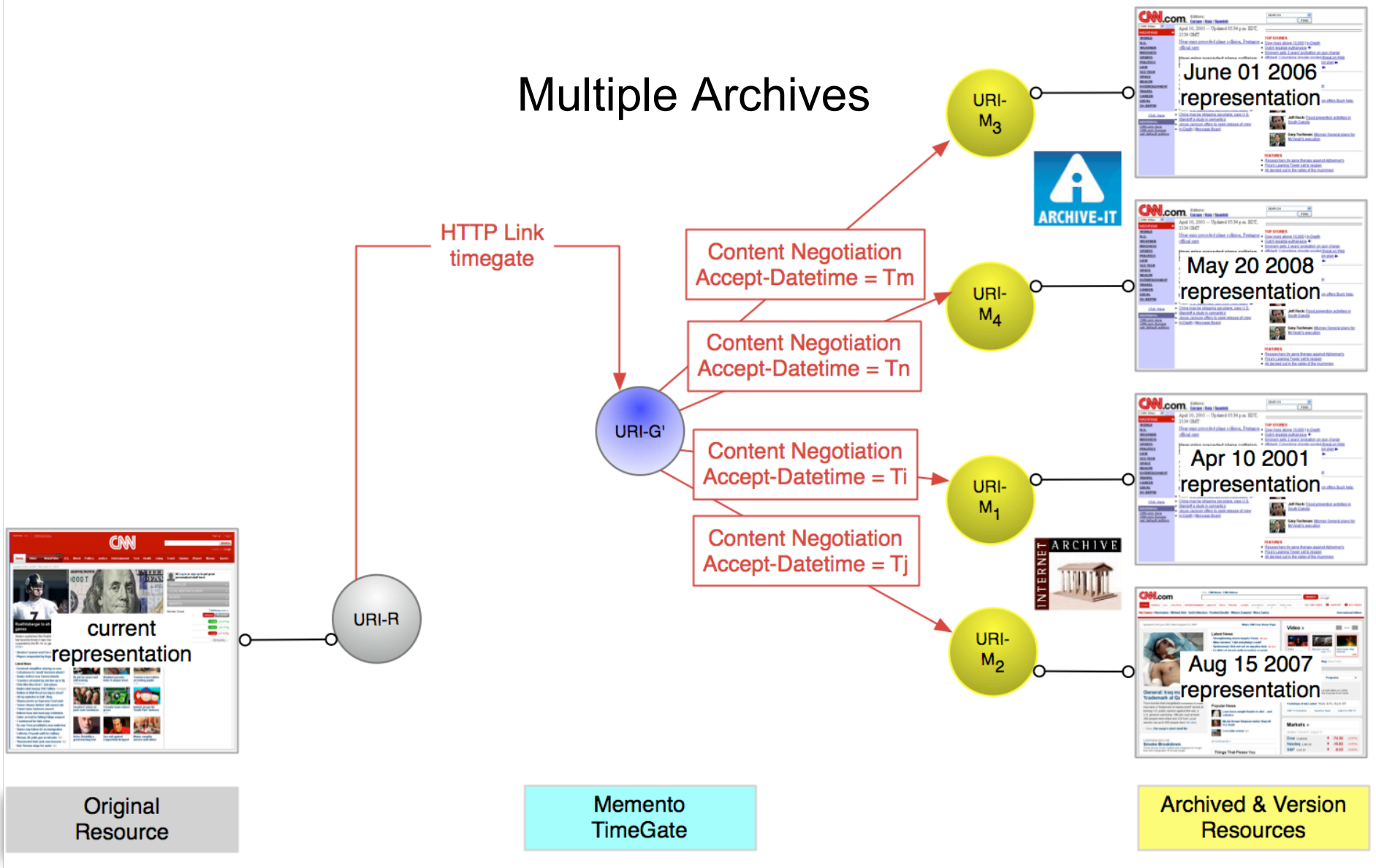


Memento
SWIB 2011, Hamburg, Germany, November 2011

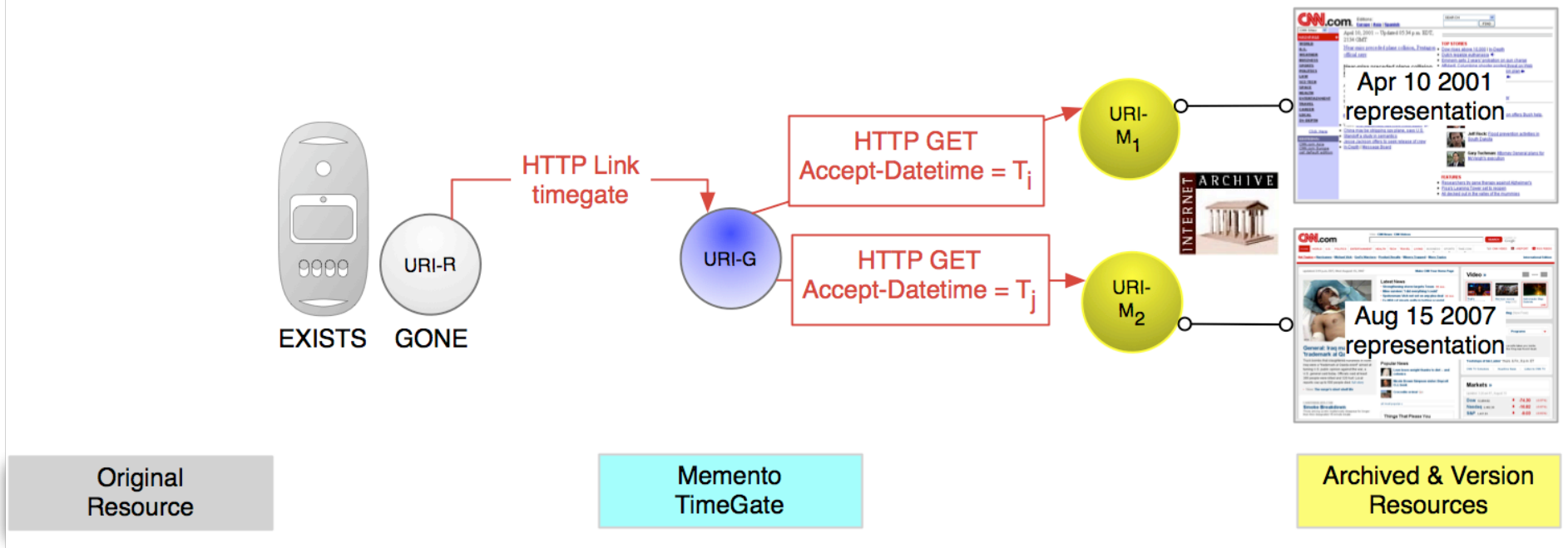
37



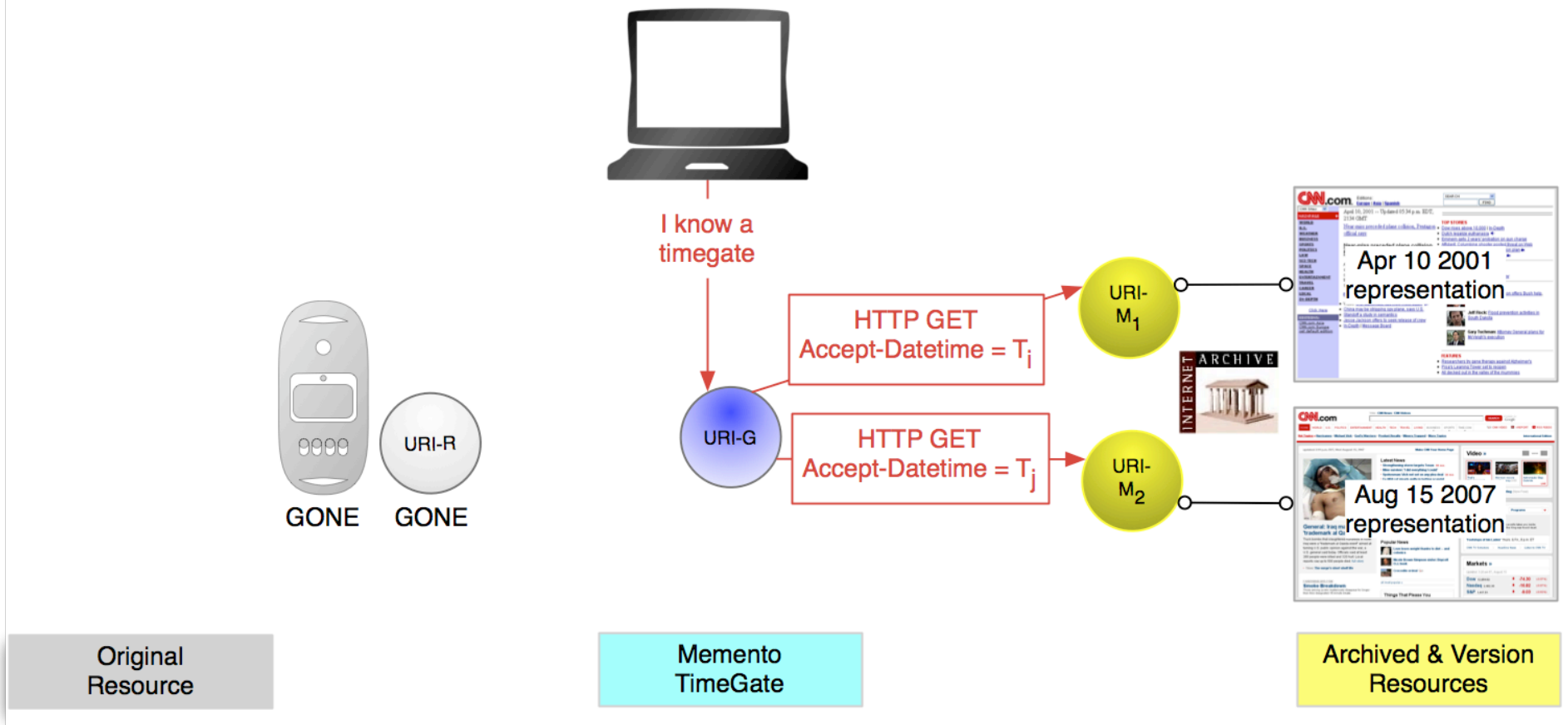
Multiple Archives



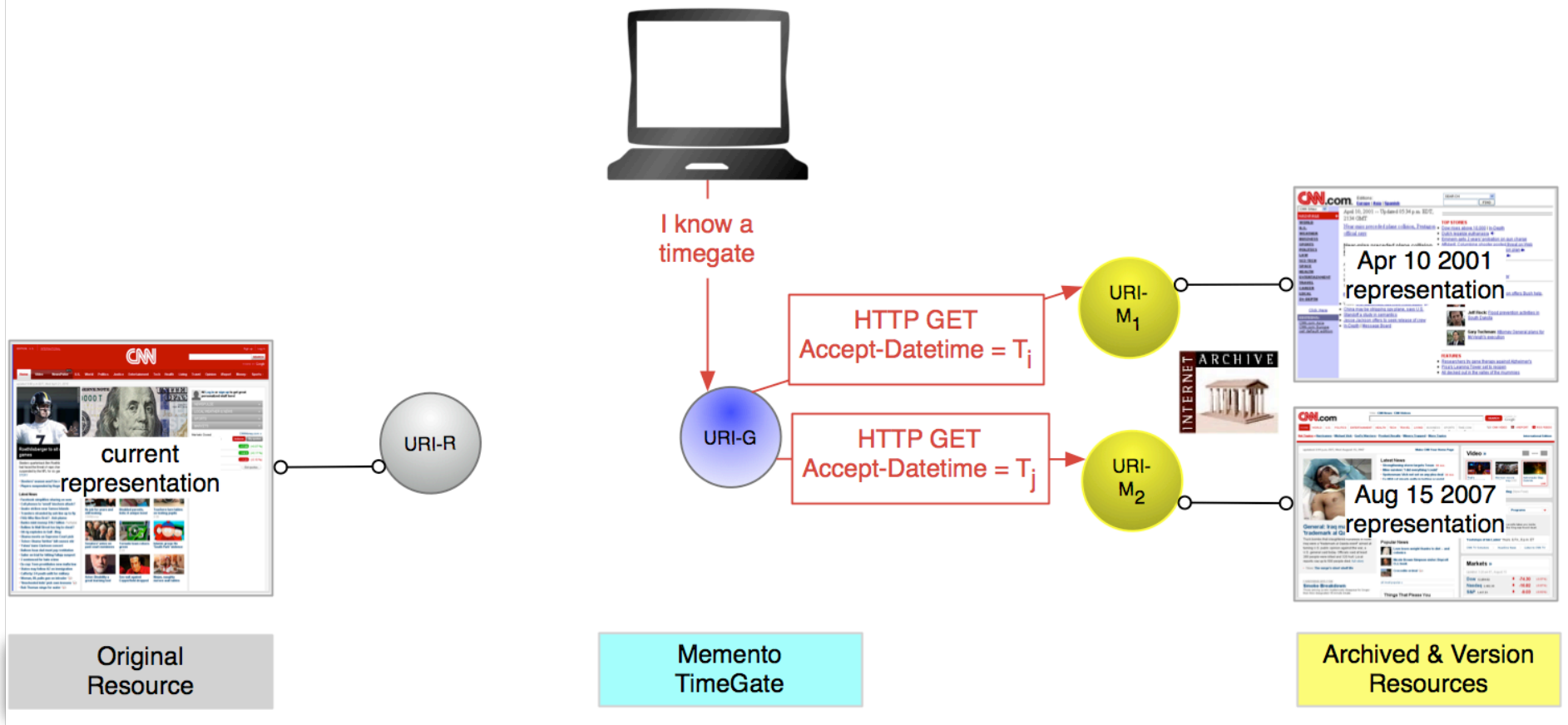
Original Resource Gone



Original Resource's Server Gone



Original Resource Provides no Link



The Memento Framework:

Protocol to Integrate Past and Current Web

HTTP Headers



HTTP Headers used in Memento

- Defines two new headers:
 - request: `Accept-Datetime`
 - response: `Memento-Datetime`
- Introduce new content for two existing headers:
 - response: `Vary ; Link`
- Use one existing headers without modification:
 - response: `Location`



HTTP Request Headers: Accept-Datetime

- Accept-Datetime
 - Issued against TimeGate, (Original Resource), (Memento)
 - Header value:
 - Desired datetime of Memento (MANDATORY)
Must be in RFC 1123 format and in GMT
 - Interval indicator to express the client is only interested in Mementos within the interval (OPTIONAL)
 - Expressed as two ISO8601 durations:
"-P3DT5H;+P2DT6H"

Accept-Datetime: Mon, 12 Oct 2009 14:20:33 GMT



HTTP Response Headers: Memento-Datetime

- Memento-Datetime
 - Returned by Mementos
 - Always. Even when not via a TimeGate
 - Header value: Archival datetime of the Memento
 - Resource has not and will not change beyond that date
 - This header is sticky:
 - Once returned, must always return it with same value
 - Must also be preserved when Mementos are mirrored at different URIs
- This header is crucial to allow a client to understand it has arrived at a Memento

See: <http://www.mementoweb.org/guide/resourcetype/>

Memento-Datetime: Mon, 12 Oct 2009 14:20:33 GMT



HTTP Response Headers: Vary

- Vary
 - Returned by TimeGate
 - Similar to regular content negotiation
 - Header value:
 - negotiate, accept-datetime
- TimeGate must first meet the datetime preference, and then – if possible – other content negotiation preferences
- Note: accept-datetime value in Vary header is crucial to allow a client to understand it has arrived at a TimeGate.
See: <http://www.mementoweb.org/guide/resourcetype/>

Vary: negotiate, accept-datetime



HTTP Response Headers: Location

- Location
 - Returned by TimeGate
 - Similar to regular content negotiation
 - Header value: URI of selected Memento

Location: `http://web.archive.org/web/20010911223004/
http://cnn.com`



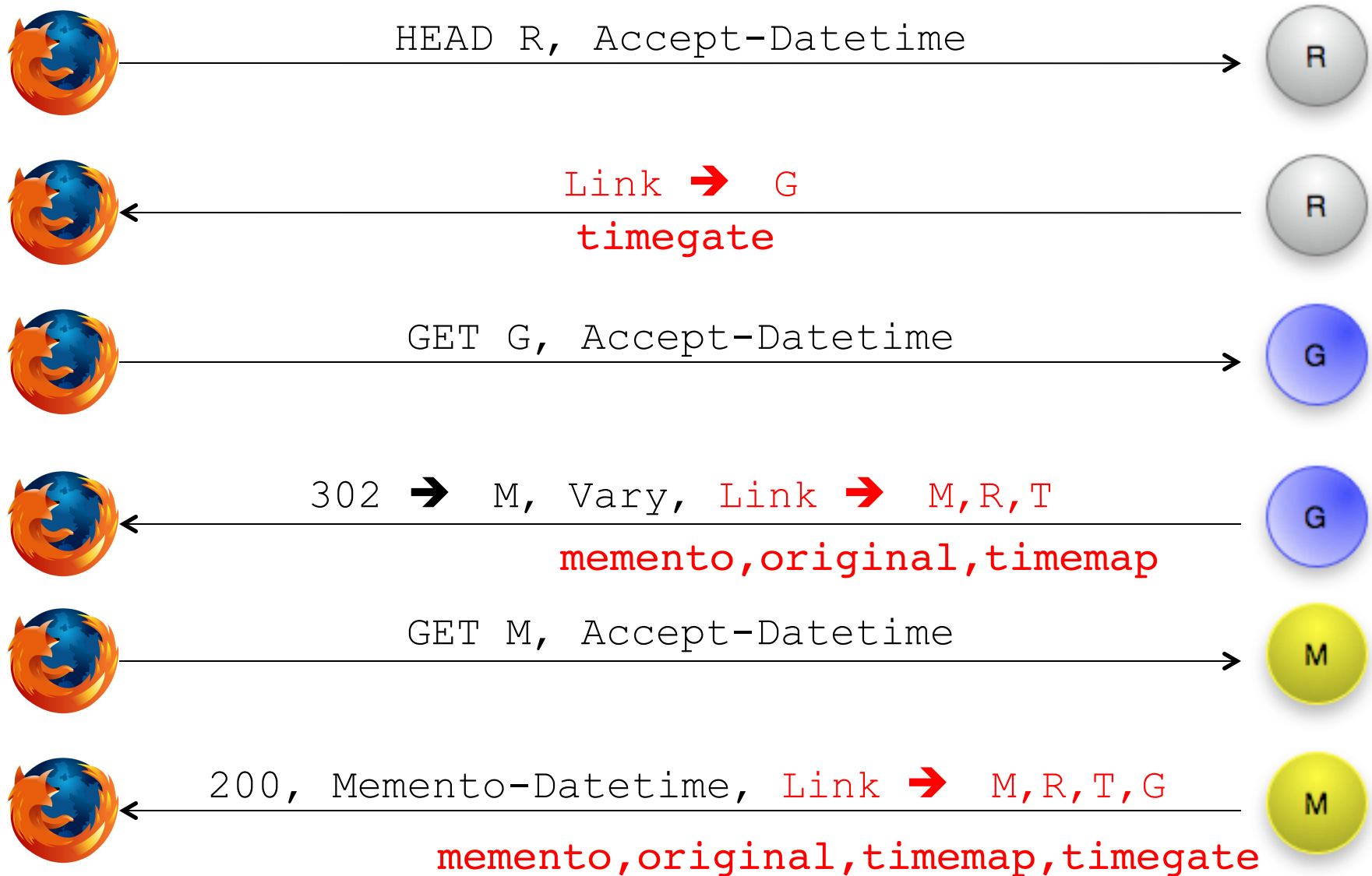
HTTP Response Headers: Link

- Link
 - Returned by Original Resource, TimeGate and Mementos
 - Various new Relation Types are introduced:
 - "original"
 - "timegate"
 - "memento"
 - "timemap"
 - Rules about usage of these Links for Original Resources, TimeGates, and Mementos
 - HTTP Link Header: RFC 5988
See: <https://datatracker.ietf.org/doc/rfc5988/>

```
Link: <http://web.archive.org/web/20010911223004/http://  
cnn.com>;rel="memento";datetime="Mon, 11 Sep 2001 22:30:04 GMT"
```



Memento HTTP Flow



The Memento Framework:

Protocol to Integrate Past and Current Web

HTTP Interactions



Memento
SWIB 2011, Hamburg, Germany, November 2011

50



Memento HTTP Flow: Step 1



```
HEAD / HTTP/1.1
Host: a.example.org
Accept-Datetime: Tue, 11 Sep 2001 20:35:00 GMT
Connection: close
```

Memento HTTP Flow: Step 2



```
HTTP/1.1 200 OK
Date: Thu, 21 Jan 2010 00:02:12 GMT
Server: Apache
Link: <http://arxiv.example.net/timegate/http://a.example.org>
      ; rel="timegate"
Content-Length: 255
Connection: close
Content-Type: text/html; charset=iso-8859-1
```

Memento HTTP Flow: Step 3



```
GET /timegate/http://a.example.org
HTTP/1.1
Host: arxiv.example.net
Accept-Datetime: Tue, 11 Sep 2001 20:35:00 GMT
Connection: close
```

Memento HTTP Flow: Step 4



```
HTTP/1.1 302 Found
Date: Thu, 21 Jan 2010 00:06:50 GMT
Server: Apache
Vary: negotiate, accept-datetime
Location:
  http://arxiv.example.net/web/20010911203610/http://a.example.org
Link: <http://a.example.org>; rel="original",
  <http://arxiv.example.net/web/20000915112826/http://a.example.org>
    ; rel="first memento"; datetime="Tue, 15 Sep 2000 11:28:26 GMT",
  <http://arxiv.example.net/web/20080708093433/http://a.example.org>
    ; rel="last memento"; datetime="Tue, 08 Jul 2008 09:34:33 GMT",
  <http://arxiv.example.net/timemap/http://a.example.org>
    ; rel="timemap"; type="application/link-format",
  <http://arxiv.example.net/web/20010911203610/http://a.example.org>
    ; rel="memento"; datetime="Tue, 11 Sep 2001 20:36:10 GMT",
  <http://arxiv.example.net/web/20010911203610/http://a.example.org>
    ; rel="prev memento"; datetime="Tue, 11 Sep 2001 20:30:51 GMT",
  <http://arxiv.example.net/web/20010911203610/http://a.example.org>
    ; rel="next memento"; datetime="Tue, 11 Sep 2001 20:47:33 GMT"
Content-Length: 0
Content-Type: text/plain; charset=UTF-8
Connection: close
```

Memento HTTP Flow: Step 5



```
GET /web/20010911203610/http://a.example.org
HTTP/1.1
Host: arxiv.example.net
Accept-Datetime: Tue, 11 Sep 2001 20:35:00 GMT
Connection: close
```

Memento HTTP Flow: Step 6



```
HTTP/1.1 200 OK
Date: Thu, 21 Jan 2010 00:09:40 GMT
Server: Apache-Coyote/1.1
Memento-Datetime: Tue, 11 Sep 2001 20:36:10 GMT
Link: <http://a.example.org>; rel="original",
      <http://arxiv.example.net/web/20000915112826/http://a.example.org>
      ; rel="first memento"; datetime="Tue, 15 Sep 2000 11:28:26 GMT",
      <http://arxiv.example.net/web/20080708093433/http://a.example.org>
      ; rel="last memento"; datetime="Tue, 08 Jul 2008 09:34:33 GMT",
      <http://arxiv.example.net/timemap/http://a.example.org>
      ; rel="timemap"; type="application/link-format",
      <http://arxiv.example.net/timegate/http://a.example.org>
      ; rel="timegate",
      <http://arxiv.example.net/web/20010911203610/http://a.example.org>
      ; rel="memento"; datetime="Tue, 11 Sep 2001 20:36:10 GMT",
      <http://arxiv.example.net/web/20010911203610/http://a.example.org>
      ; rel="prev memento"; datetime="Tue, 11 Sep 2001 20:30:51 GMT",
      <http://arxiv.example.net/web/20010911203610/http://a.example.org>
      ; rel="next memento"; datetime="Tue, 11 Sep 2001 20:47:33 GMT"
Content-Length: 23364
Content-Type: text/html;charset=utf-8
Connection: close
```


The Memento Framework:

Discovery to Support Integration of Past and Current Web



Memento
SWIB 2011, Hamburg, Germany, November 2011 ⁵⁷



Discovery to the Memento Framework

- Important:
 - To support discovery of TimeGates when servers of Original Resources do not provide “timegate” Links (which is currently the case for 99.99999 % of Web servers);
 - To allow emergence of aggregating TimeGates that cache metadata about Mementos in multiple archives (avoiding performance issues caused by on-the-fly negotiating with multiple remote TimeGates).
- Approaches to support discovery of:
 - TimeGates;
 - TimeMaps (and hence Mementos and TimeGates);
 - Mementos.



The Memento Framework:

Discovery to Support Integration of Past and Current Web

TimeGate Discovery



Memento
SWIB 2011, Hamburg, Germany, November 2011

59



Batch Discovery of TimeGates: robots.txt

- robots.txt file is used by Web servers to convey crawling policies
- Web crawlers (such as for archives) retrieve and parse it
- De-facto standard, no official endorsement
- Extended with new directives, including by Google

```
User-agent: *           # Reject all crawlers
Disallow: /

                        # Google Sitemap extension
Sitemap: http://some.example.com/me/sitemap.xml

User-agent: NiceBot    # Select only NiceBot
Crawl-delay: 10        # 10 seconds between requests
```

Investigating Well-Known URI <http://tools.ietf.org/html/rfc5785>



Batch Discovery of TimeGates: robots.txt

- Add TimeGate and Archived directives to support discovery of TimeGates known to the server
- User agent should concatenate desired URL with TimeGate link
- Archived value is truncated host/path or * to describe a general web archive

```
TimeGate: http://a.example.org/w/index.php/Special:TimeGate/  
Archived: a.example.org/w/
```

```
TimeGate: http://arxiv.example.net/timegate/  
Archived: a.example.org/  
Archived: www.a.example.org/
```

```
TimeGate: http://arxiv.example.net/timegate/  
Archived: *
```

<http://mementoweb.org/guide/robotstxt/>



Memento

SWIB 2011, Hamburg, Germany, November 2011

61



The Memento Framework:

Discovery to Support Integration of Past and Current Web

Memento Discovery: All Mementos known by an archive



Memento
SWIB 2011, Hamburg, Germany, November 2011⁶²



Batch discovery of Mementos: robots.txt

- robots.txt file is used by Web servers to convey crawling policies
- Support discovery of Mementos through robots.txt via existing User-agent and Allow directives
- Use value memento for User-agent to convey that the value for the Memento-Datetime header must remain sticky when crawling/mirroring Mementos

```
User-agent: *  
Disallow: /  
User-agent: memento  
Allow: /web/
```



Batch discovery of Mementos: Memento Feeds

- Concept:
 - Archives publish feeds in which each entry provides details about a specific Memento, e.g. Memento-Datetime, Original Resource, etc.
 - As new Mementos become available, new feeds with new entries are published
 - Once published, feeds remain static
- Technology:
 - To be decided in collaboration with IIPC
 - Inspired by the approach and functionality of CDX files (see http://www.archive.org/web/researcher/cdx_legend.php) but:
 - With Memento-specific extensions;
 - Possibly using different serialization;
 - Including mechanisms to discover these feeds.



The Memento Framework:

Discovery to Support Integration of Present and Past Web

**Discovery via TimeMaps: All Mementos for a given
Original Resource known by an archive**



Memento
SWIB 2011, Hamburg, Germany, November 2011

65



TimeMap Overview

- A TimeMap is an inventory of Mementos for an Original Resource that the responding server is aware of. It lists at least:
 - URI of Original Resource
 - URI and datetime of all known Mementos
 - URI of TimeGate for Original Resource
 - URI of TimeMap itself
- Multiple TimeMap serializations possible:
 - application/link-format mandatory

see *<https://datatracker.ietf.org/doc/draft-ietf-core-link-format/>*

- RDF TimeMaps proposed



TimeMaps: Link Format Syntax

- Document in the format of the value of the Link HTTP Header
- Format:
 - `<URI>;rel="type";attr="val",`
 - `<URI2>...`
- `rel` is the relationship between context URI and the URI in `<>s`
- The Context URI for TimeMaps is the URI with `rel="original"`
- Other `rel` types link to Mementos, TimeGates etc.

```
<http://cnn.com/>;rel="original",  
<http://web.archive.org/web/20010911223004/http://  
cnn.com>;rel="memento";datetime="Mon, 11 Sep 2001 22:30:04 GMT"
```



TimeMaps: Link Attributes

- Memento rel types:
 - "original" The Original Resource
 - "memento" A Memento of the Original
 - "timegate" A TimeGate for the Original
 - "timemap" A TimeMap of Mementos of the Original
- Existing Attributes for Links
 - "rel" The type of relationship
 - "type" The (mime) format of the linked resource
 - "title" Title of the linked resource
 - "hreflang" Language of linked resource
 - "media" Intended media (eg screen)
 - "anchor" URI to override context URI for link



TimeMaps: Link Attributes

- New Attribute for Mementos:
 - `datetime` The Memento-Datetime
- Proposed new Attributes for Mementos:
 - `license` License associated with Memento
 - `embargo` Time after which Memento is available

```
<http://cnn.com/>;rel="original",  
<http://web.archive.org/timegate/cnn.com/>;rel="timegate",  
<http://web.archive.org/timemap/link/cnn.com/>;rel="timemap",  
<http://web.archive.org/web/20010911223004/http://  
  cnn.com>;rel="memento";datetime="Mon, 11 Sep 2001 22:30:04 GMT"  
<http://web.archive.org/web/.../cnn.com/>;rel="memento";  
  license="http://archive.org/license/1";  
  datetime="...";embargo="Mon, 20 Jul 2011 00:00:00 GMT",  
...
```



The Memento Framework:

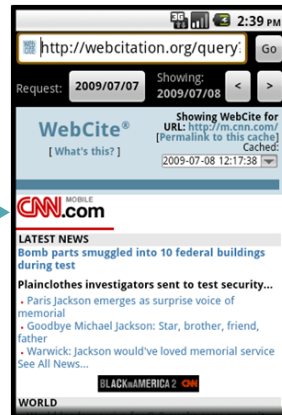
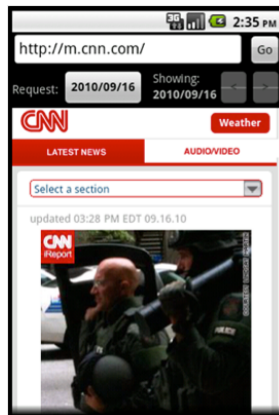
Tools



Memento
SWIB 2011, Hamburg, Germany, November 2011 70



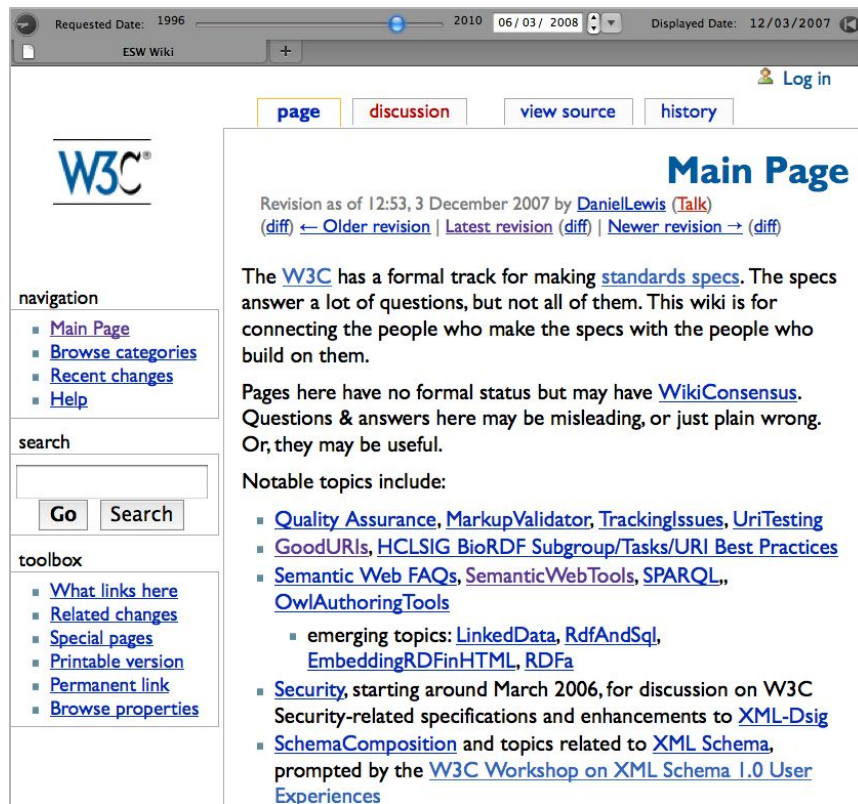
Memento Client Support



- Several client tools developed by us and others
- Add-ons for Firefox (operational) and Internet Explorer (experimental)
- Applications for Android (operational) and iPhone/iPad (in development)
- mcurl (in development)
- Paper in Code4Lib Journal
<http://journal.code4lib.org/articles/4979>



Memento Server Support



- Plug-in for MediaWiki (operational)
- Used on W3C's main wiki
- Please install it for your MediaWiki!

<http://www.mediawiki.org/wiki/Extension:Memento>

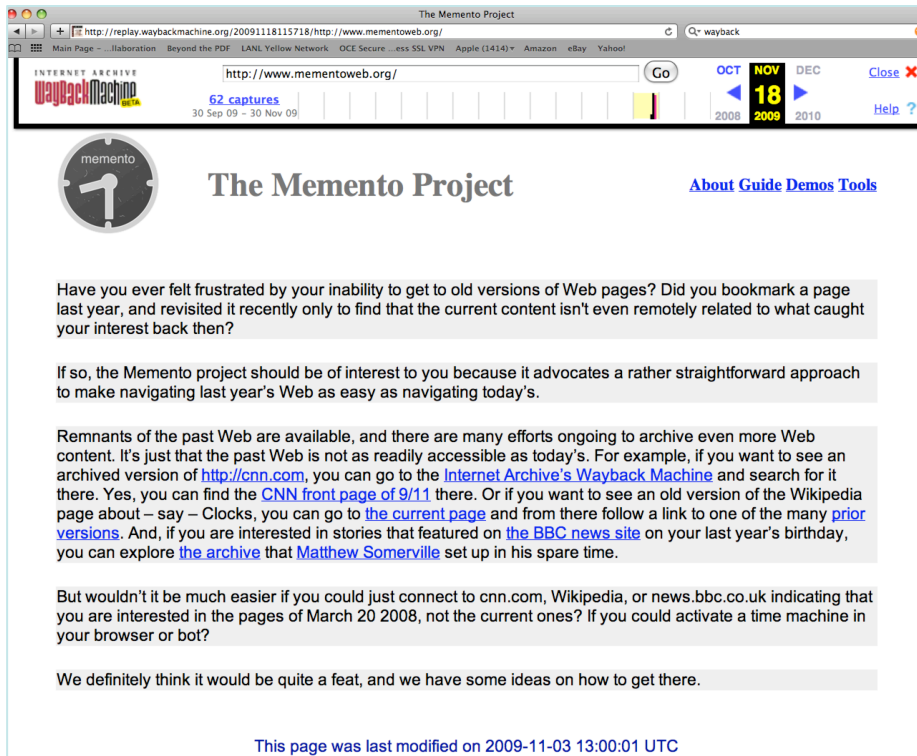


Memento
SWIB 2011, Hamburg, Germany, November 2011

72



Memento Server Support



- Memento-compliant Wayback software:
 - In production at the Internet Archive
 - Available to Web archives, worldwide
 - Please have your favorite Web Archive experiment with the new 1.6 version!



<http://mementoweb.org/tools/wayback/>



Memento
SWIB 2011, Hamburg, Germany, November 2011

73



Memento Server Validator

Test Details

Tests for TimeGate: http://memento.waybackmachine.org/memento/timegate/http://www.ietf.org/		
Test Result	URI/Date Returned	
✓	TimeGate returns status 302	
✓	TimeGate has Location header for Memento	http://memento.waybackmachine.org/memento/20100402222407/http://www.ietf.org/
✓	TimeGate acknowledges Accept-Datetime in Vary header	
✓	Link Header has link with rel=original	http://www.ietf.org/
✓	Link Header has link with rel=timebundle	http://memento.waybackmachine.org/list/timebundle/http://www.ietf.org/
✓	Link Header has link with rel=timemap	http://memento.waybackmachine.org/list/timemap/link/http://www.ietf.org/
✓	TimeMap Link has application/link-format type	
✓	Link Header has one or more links with rel=memento	http://memento.waybackmachine.org/memento/20100219084310/http://www.ietf.org/
✓	-- Selected Memento in Link Header with rel=memento	
✓	Memento has parsable datetime attribute	Fri, 19 Feb 2010 08:43:10 GMT
✓	Memento has parsable datetime attribute	Wed, 06 Nov 1996 11:49:54 GMT

- Server side client:
 - Attempts to perform all Memento actions against a given URI
 - Reports success/failure of the interactions and warnings for optional aspects
 - Kept up to date with IETF Internet Draft

<http://mementoweb.org/tools/validator/>



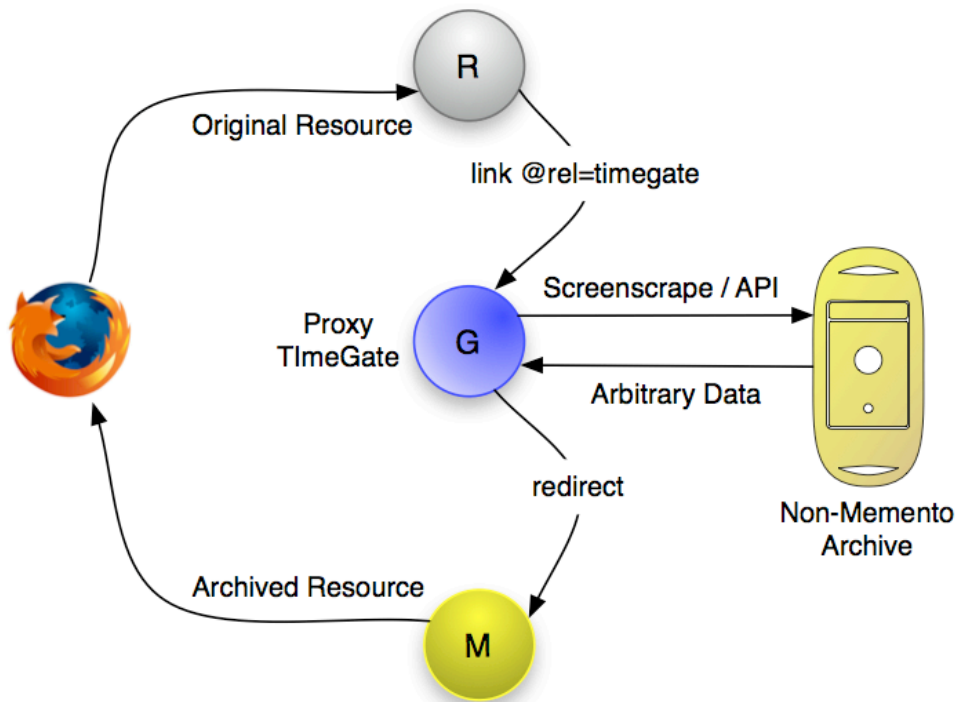
Memento
SWIB 2011, Hamburg, Germany, November 2011

74

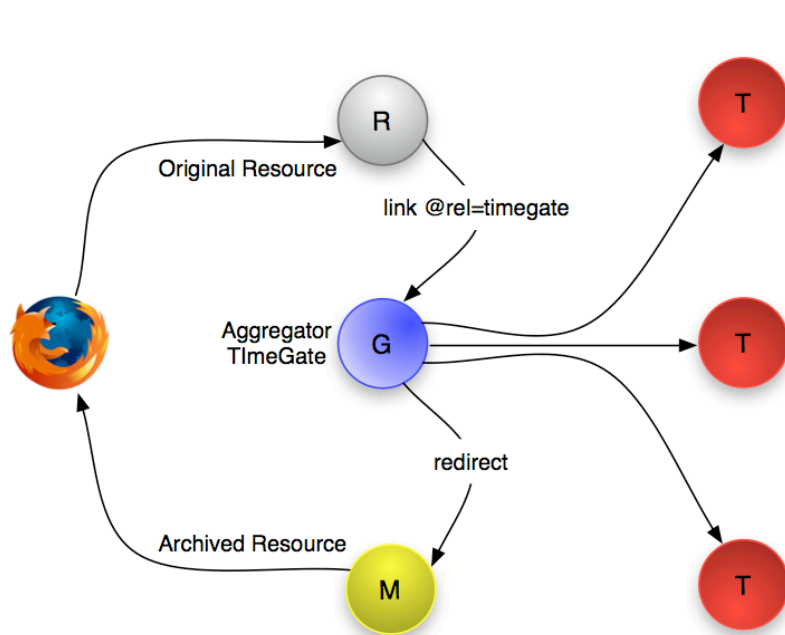


Memento Proxy Support

- Several systems that host Mementos made Memento-compliant “by proxy”
- Many major Web Archives that do not yet run Memento-compliant software
- 3,000+ MediaWiki systems, including Wikipedia, Wikia
- We would love all of these to become natively Memento compliant!



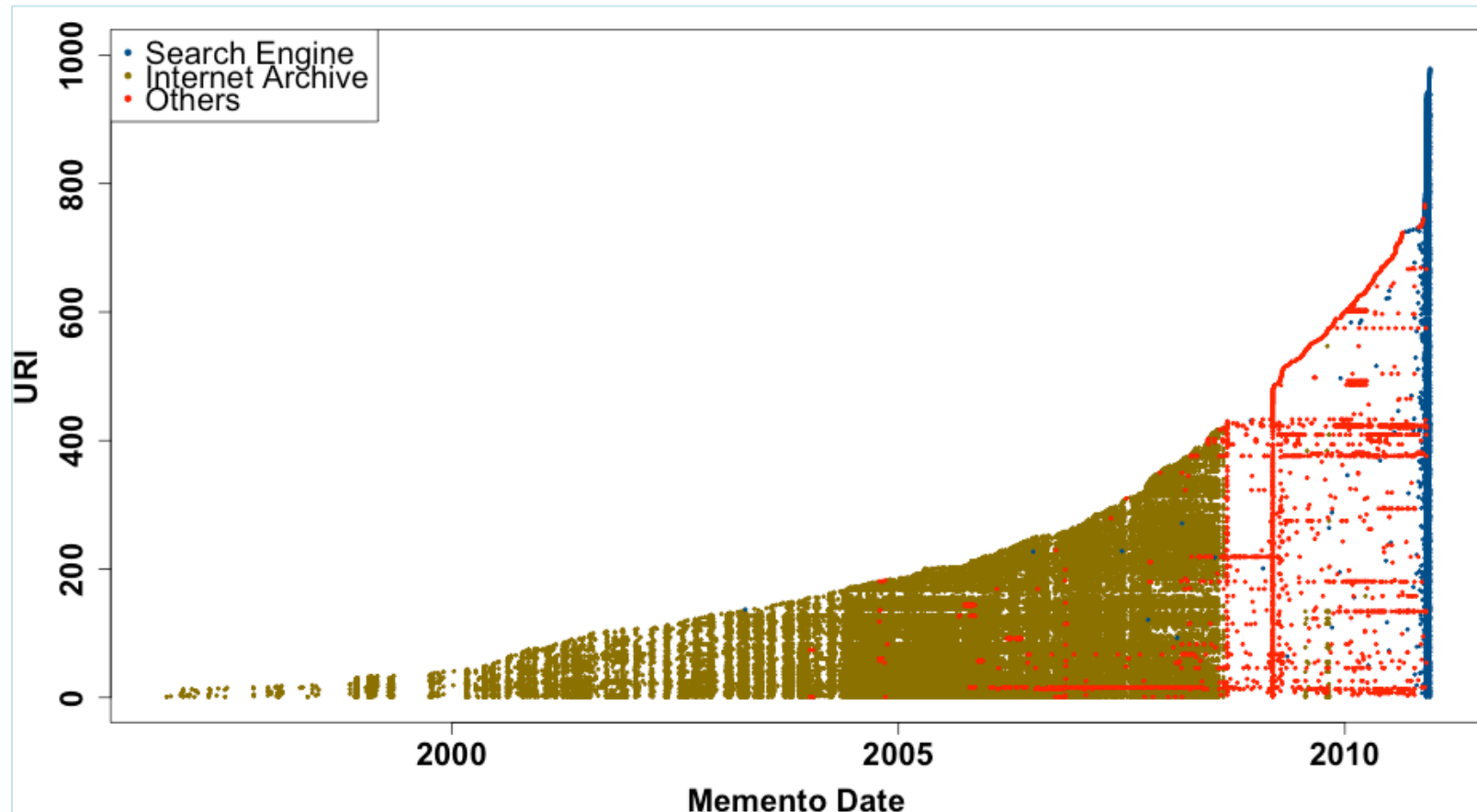
Memento Aggregator TimeGate



- Aggregates all known TimeGates
 - Proxies
 - Native Implementations
- Redirects to authoritative TimeGates (Wikipedia, Transactional Archives)
- Currently implemented with BerkeleyDB
- Future version to use FaceBook's Cassandra platform



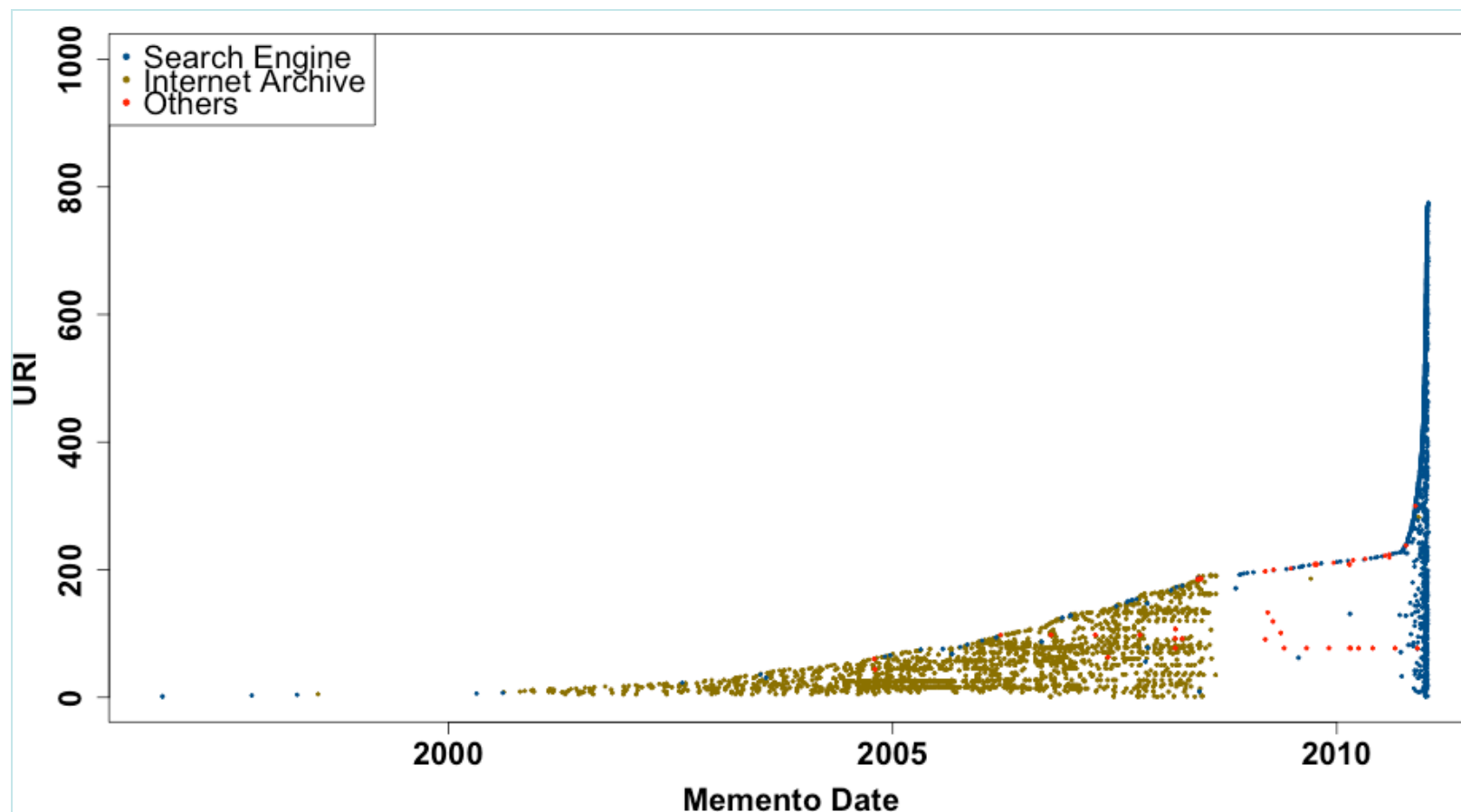
Aggregators Find More Mementos!



- 1000 URIs sampled from delicious.com
- 1 dot = 1 Memento (x=Memento-Datetime, y=URI of Original Resource)
- Sorted by URI longevity



But Still Too Few Mementos To Be Found...



- 1000 URIs sampled from search engine result pages;
- See: “How Much of the Web is Archived?” JCDL 2011



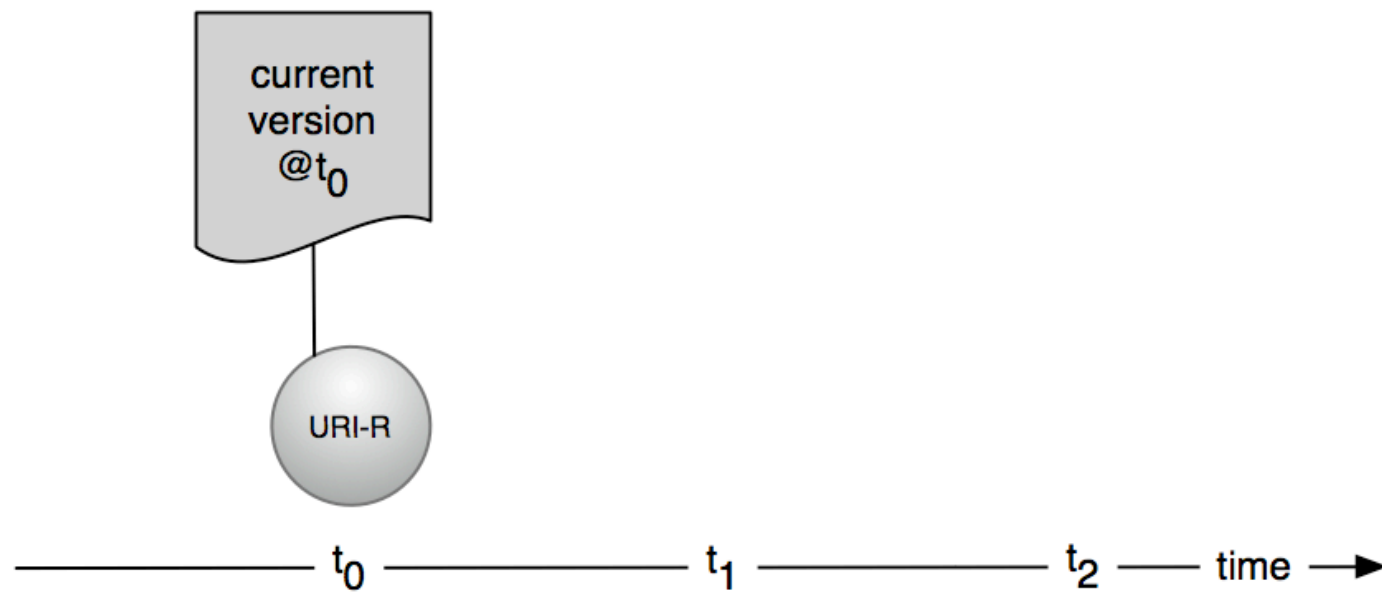
The Memento Framework:

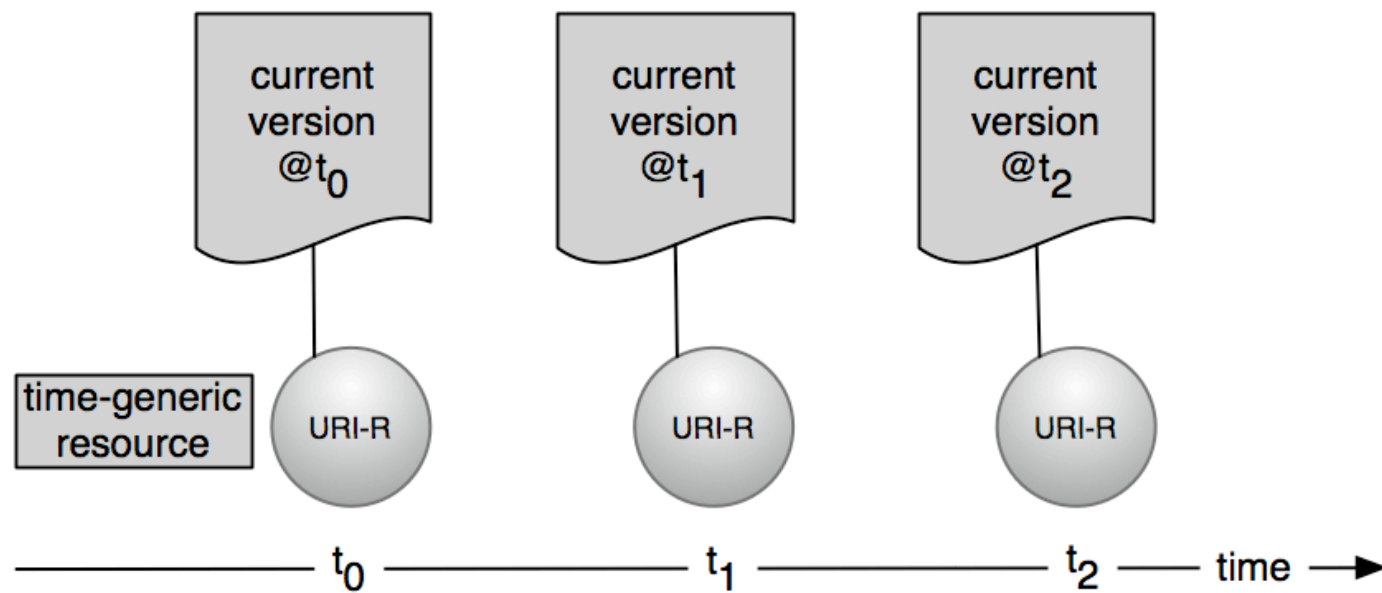
Resource Versioning

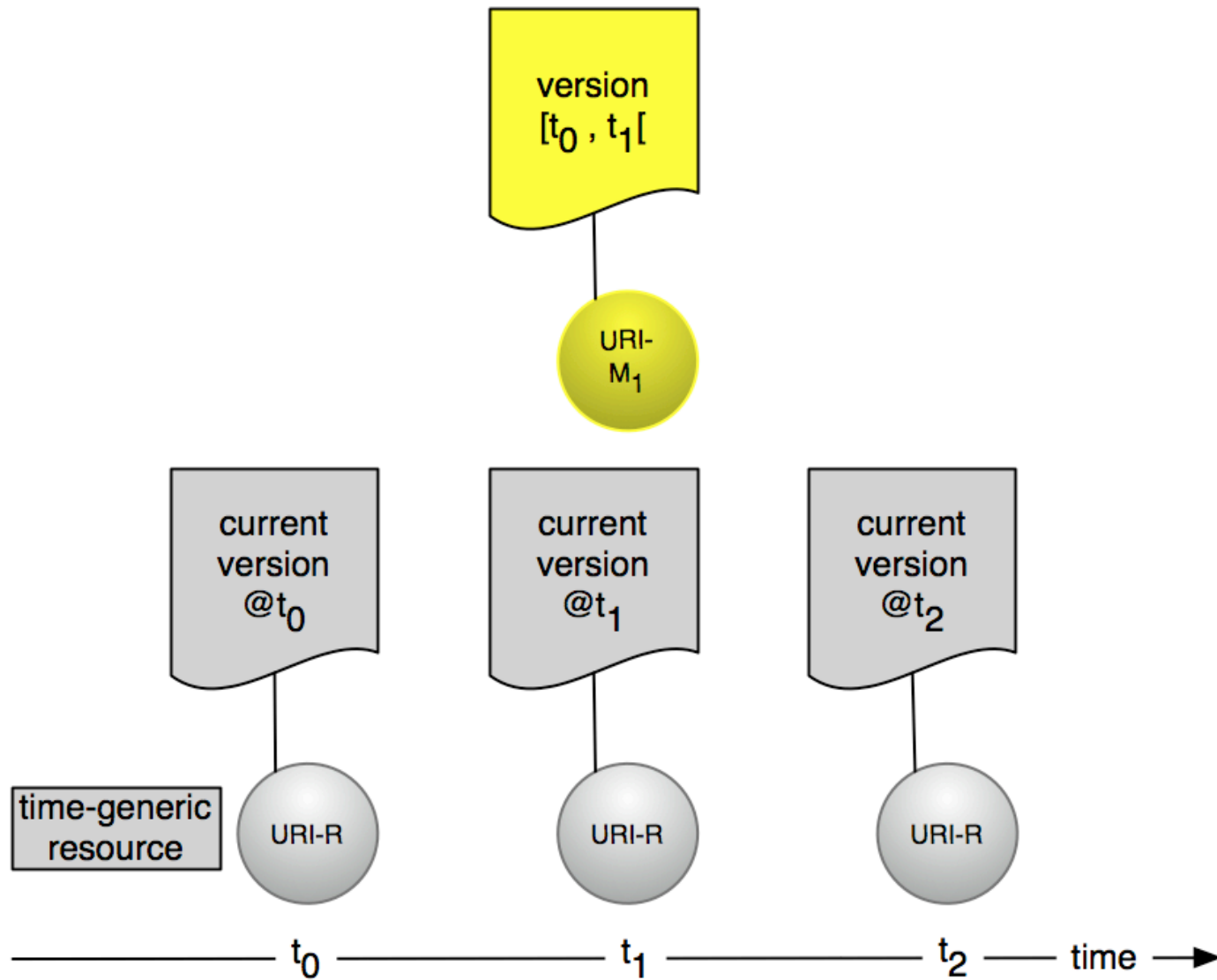


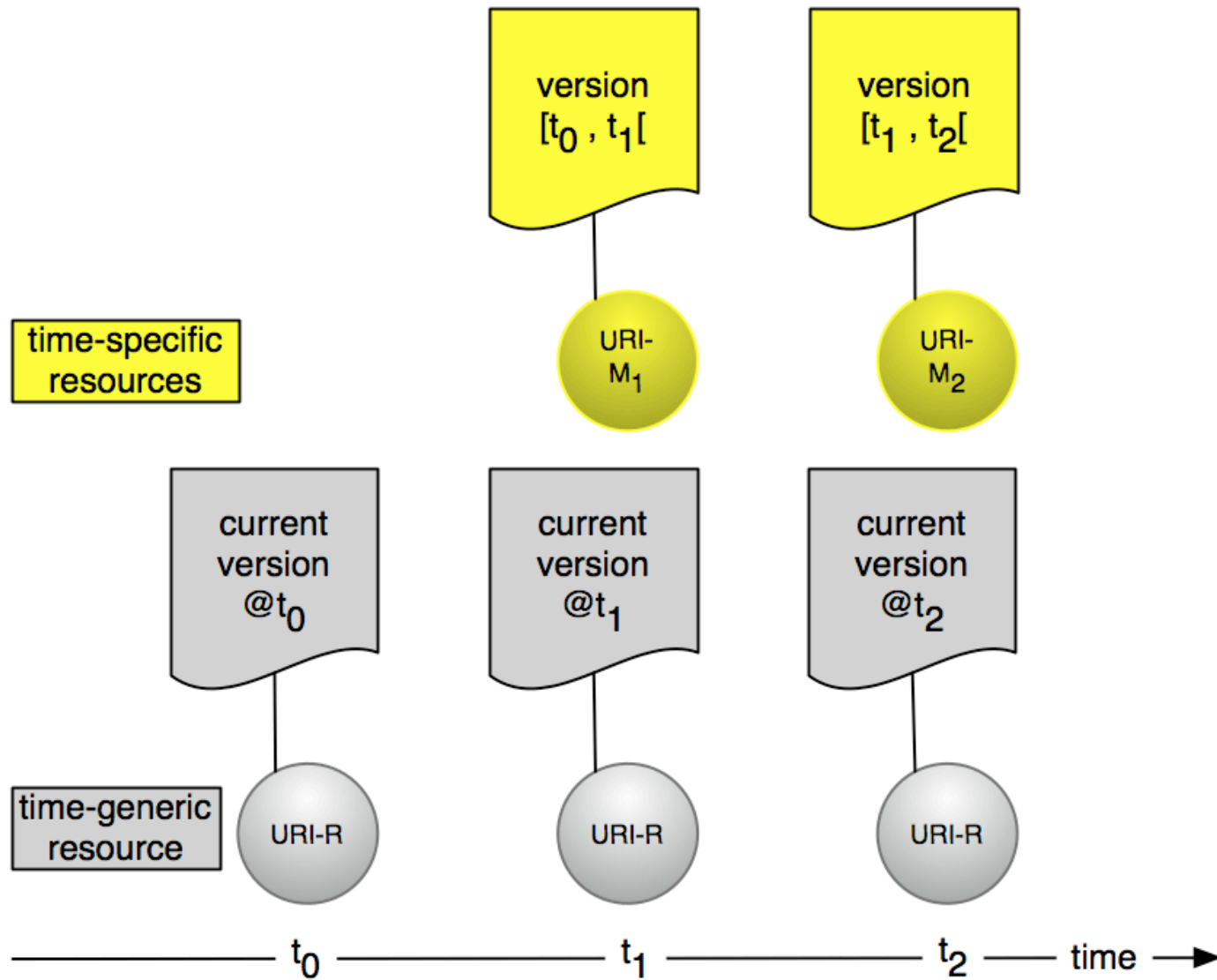
Memento
SWIB 2011, Hamburg, Germany, November 2011 ⁷⁹

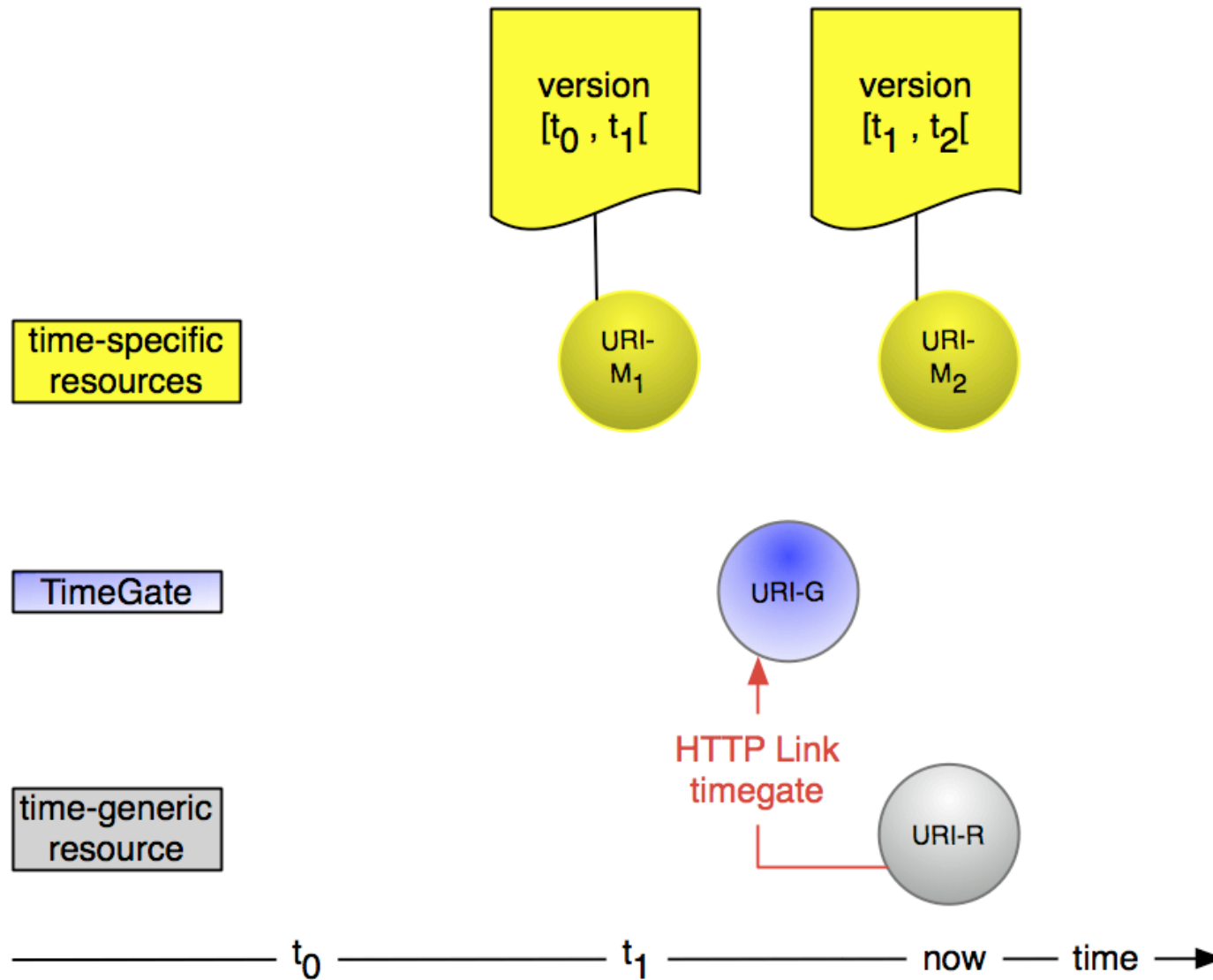


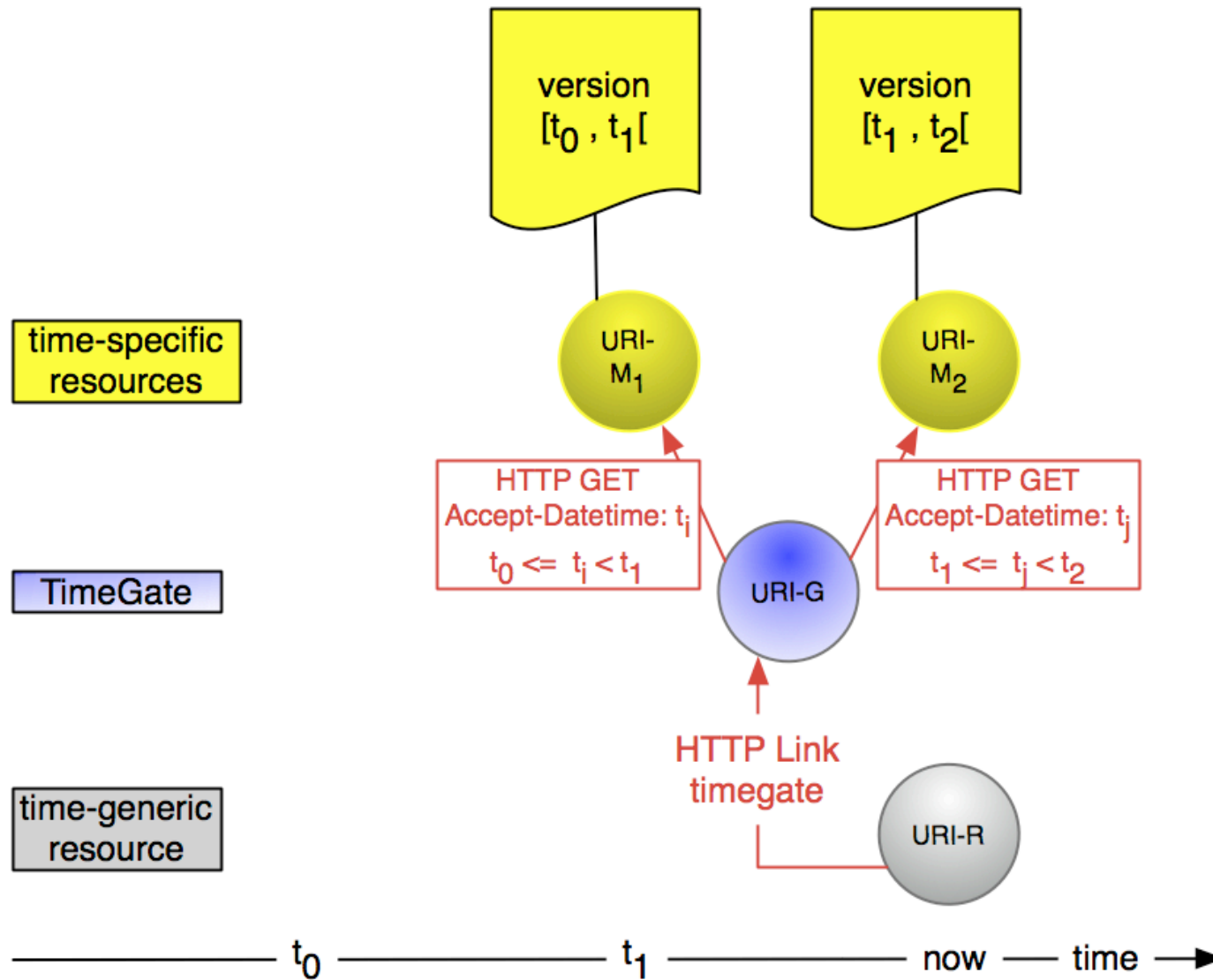


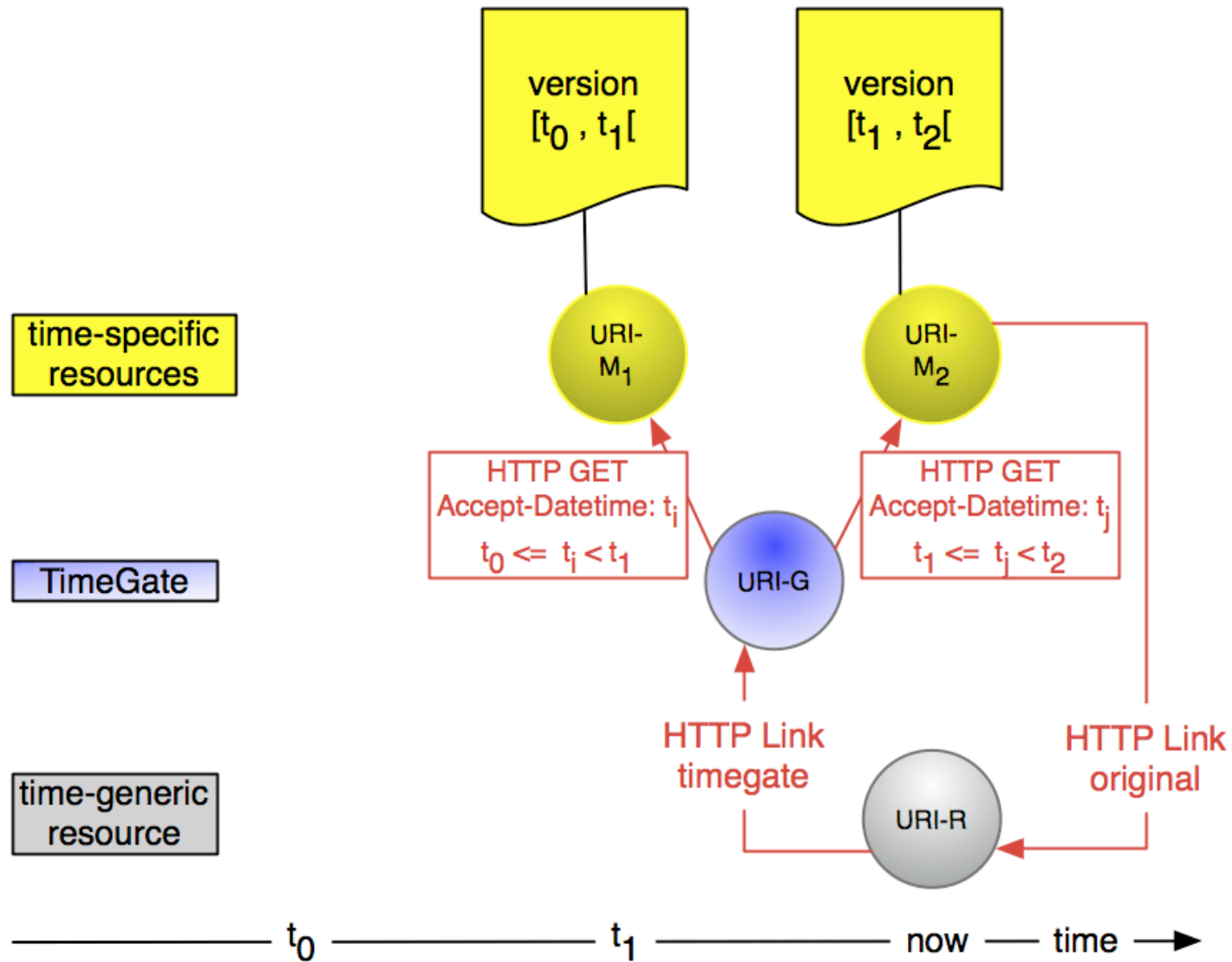






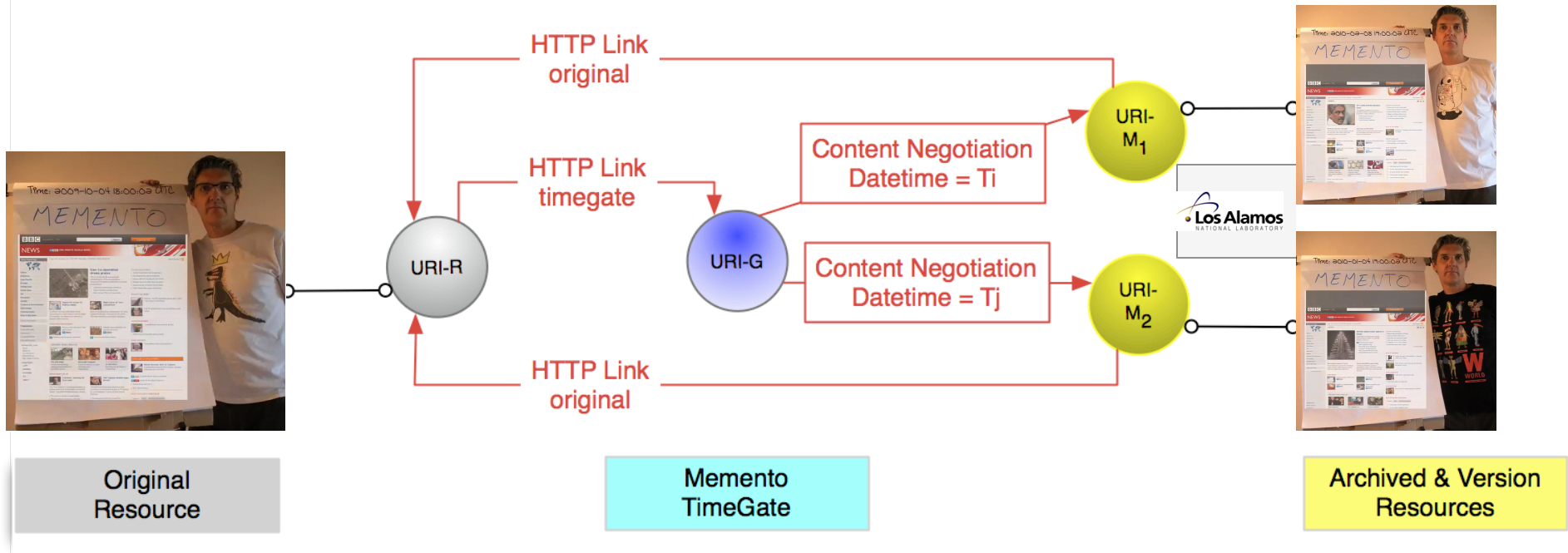






Memento Framework

Original Resource: <http://lanlsource.lanl.gov/pics/picoftheday.png>



Time Travel across Versions of a Picture of the Day



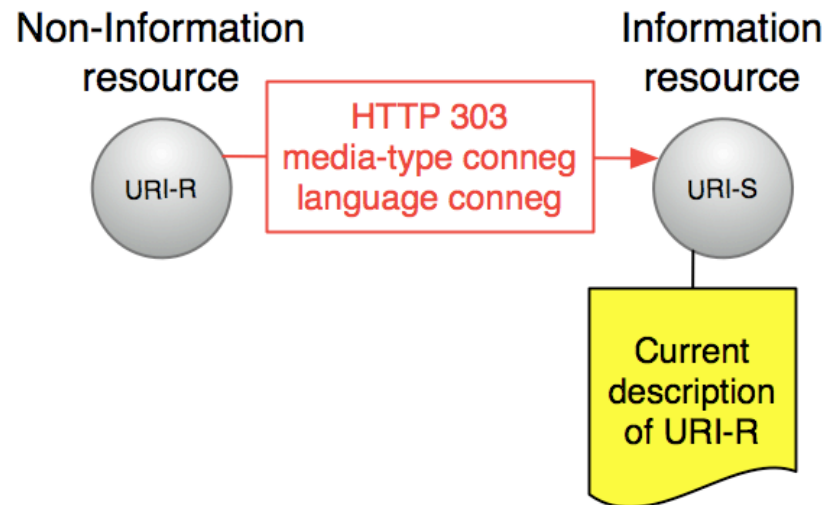
Movie at: <http://www.mementoweb.org/demo/picoftheday.mov>

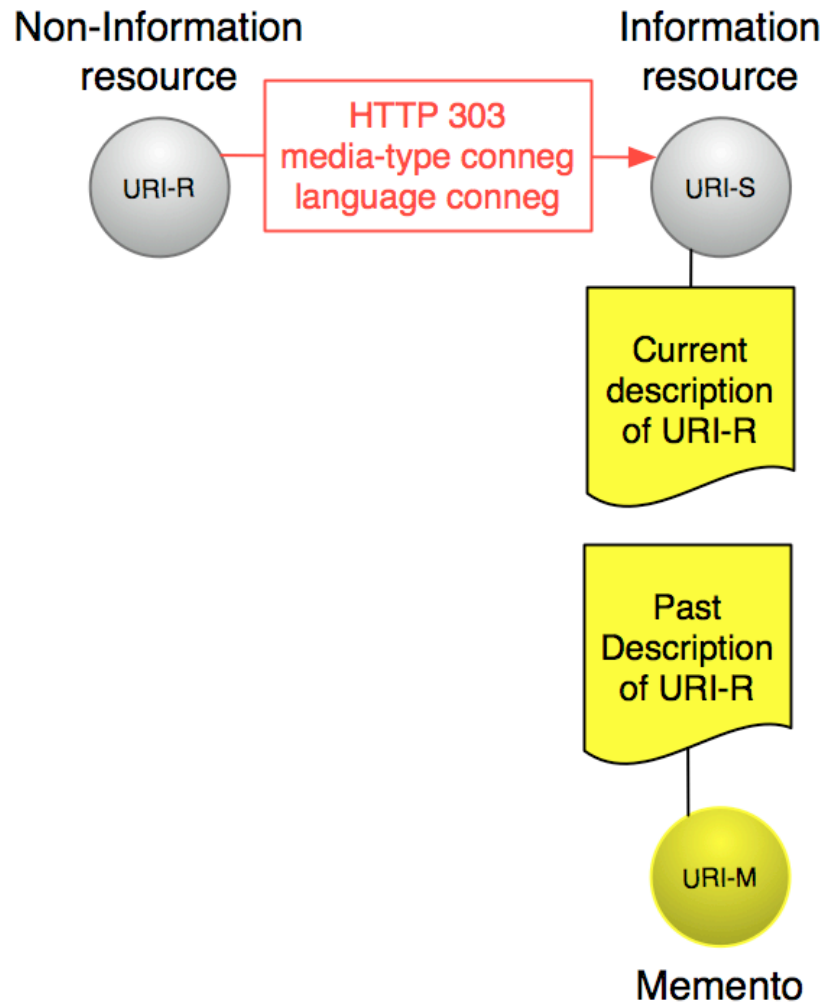


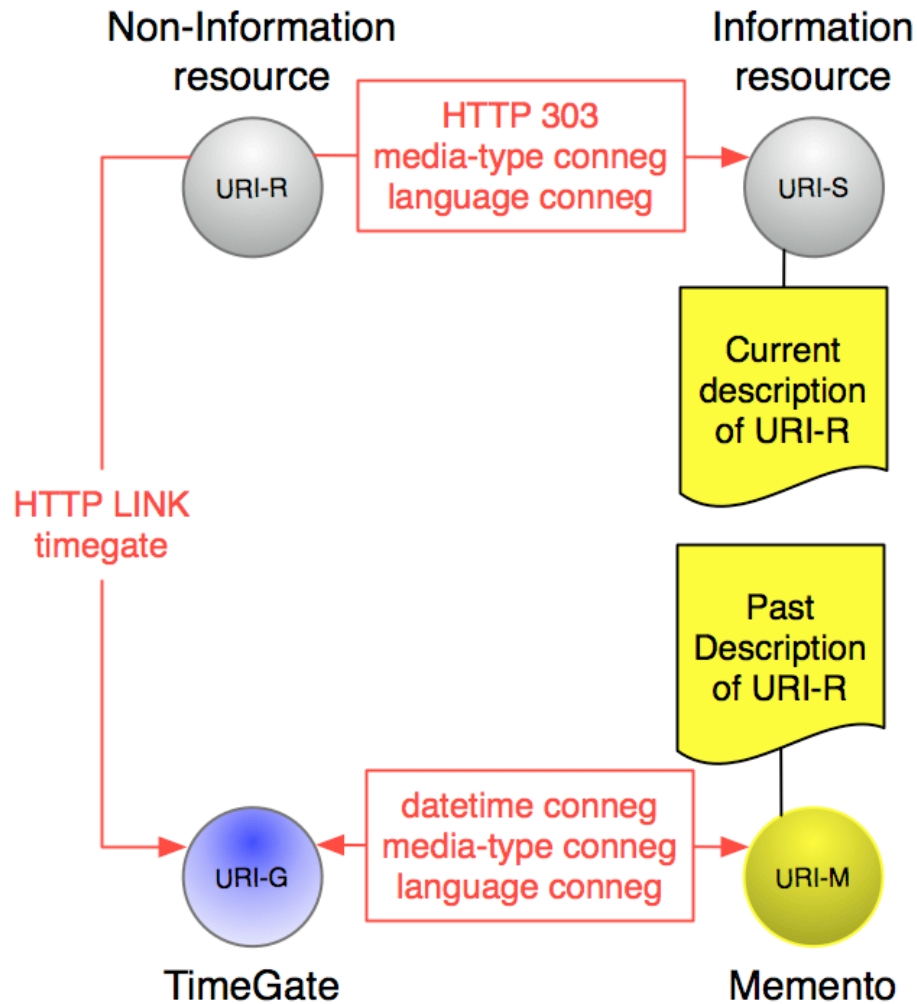
Memento
SWIB 2011, Hamburg, Germany, November 2011

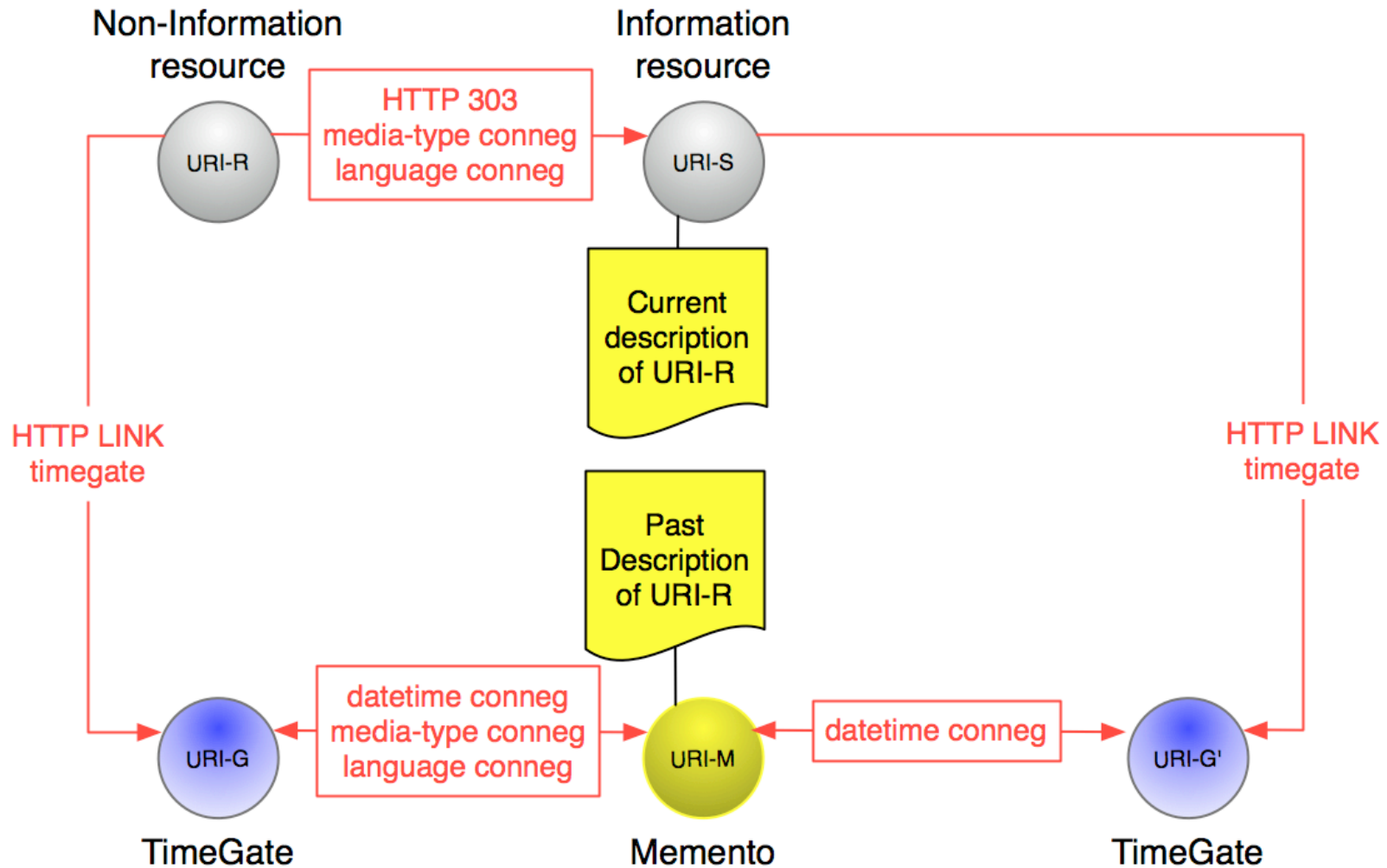
88

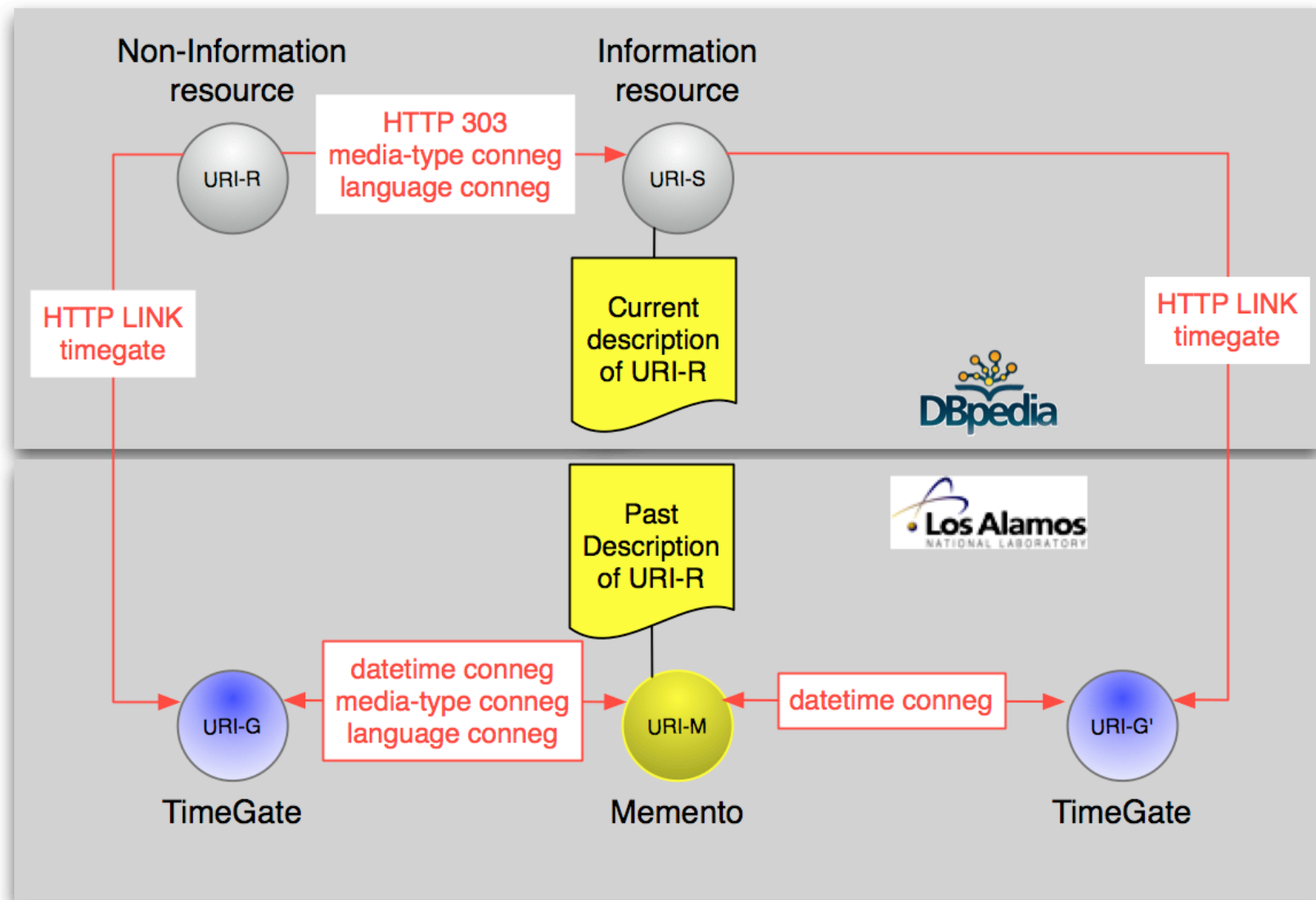






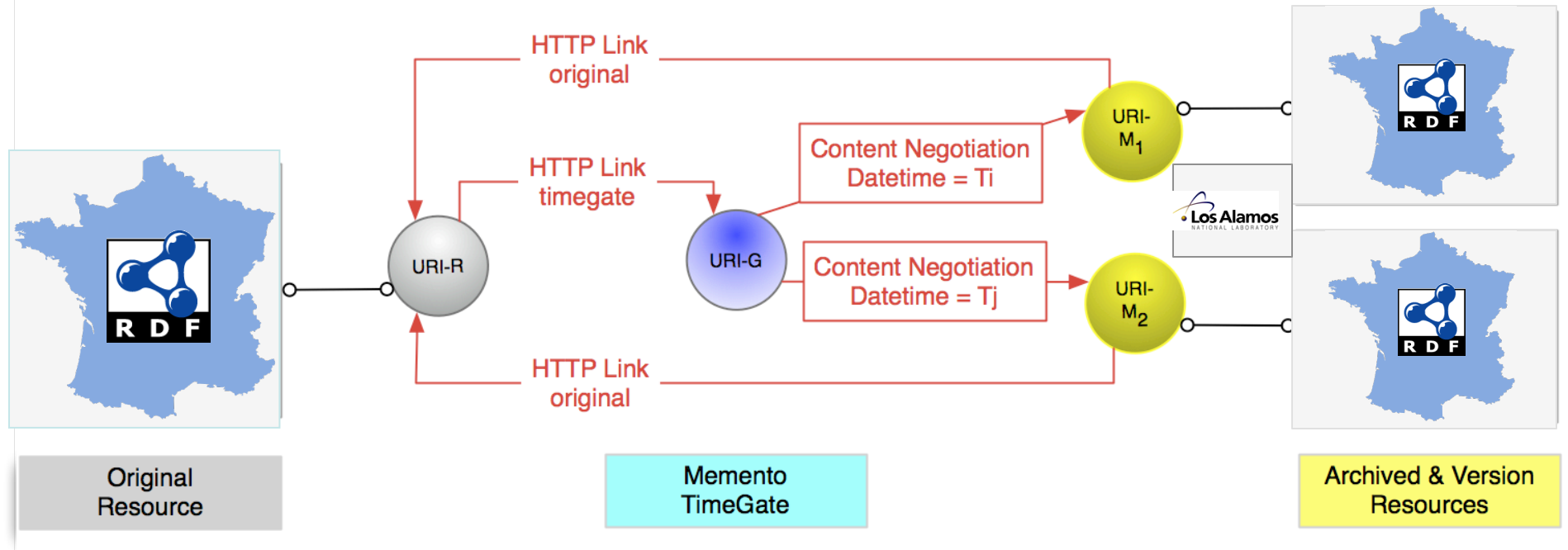




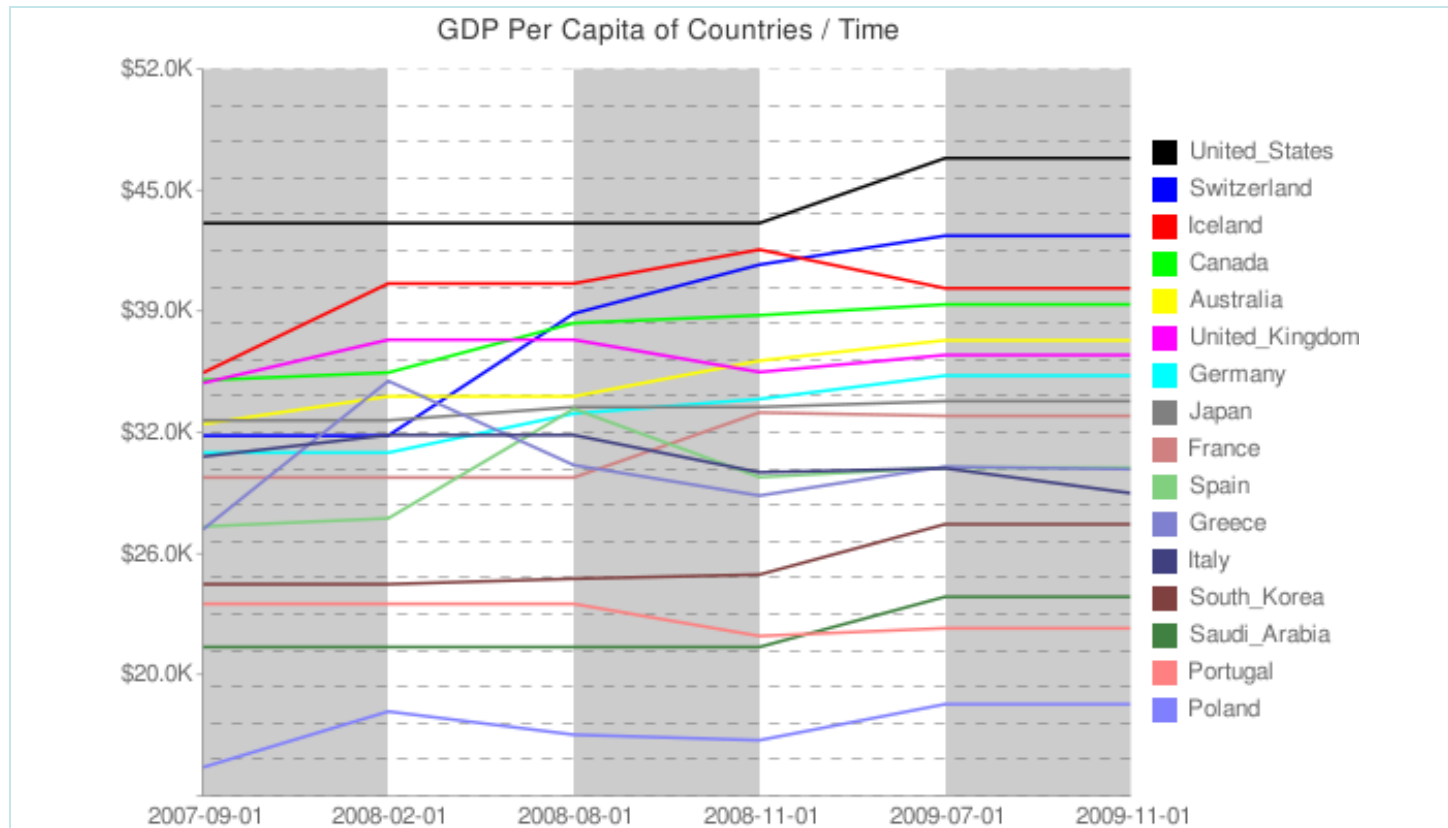


Memento Framework

Original Resource: <http://dbpedia.org/resource/France>



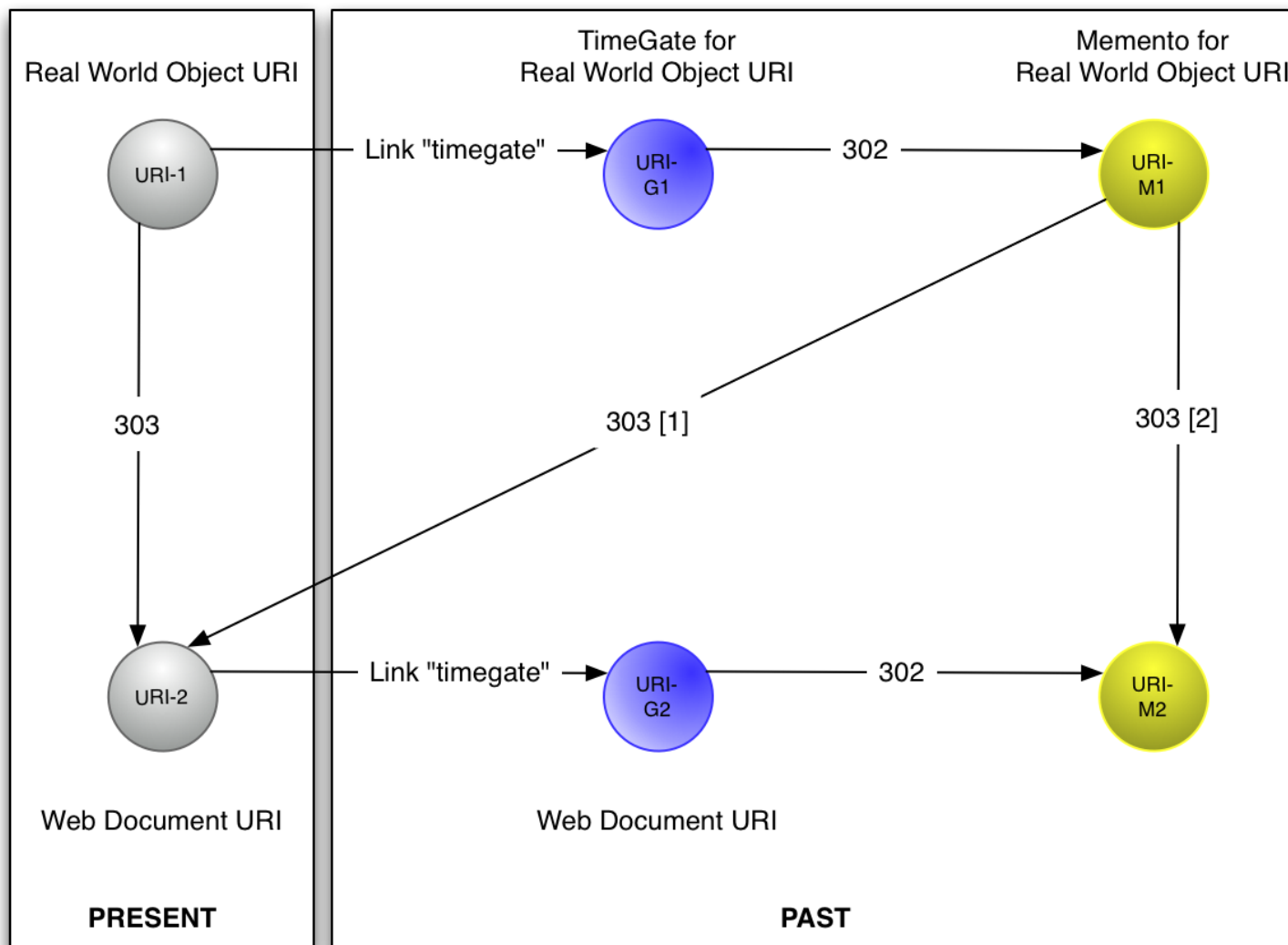
Time-Series Analysis across DBpedia Versions



Data collected through HTTP Navigation

paper at <http://arxiv.org/abs/1003.3661>





Depending on implementation of an archival solution, path [1] or [2] is available. [1] is a replay of an archived 303 that takes the client back to an information resource on the current web, where the Memento cycle can be repeated. [2] is a 303 directly to the appropriate Memento that has the description of the Real World Object for the desired point in time.



The Memento Framework:

Persistent Web Annotations



Memento
SWIB 2011, Hamburg, Germany, November 2011 ⁹⁷



Web-Centric Annotation: No Persistence

[Paul Murray](#) - Mar 8, 2010

Lead story this morning.

With a magnitude six earthquake, attacks around the world killing hundreds and a domestic civil servants strike, I find it very odd that the BBC website should have chosen to lead with "Oscar triumph for The Hurt Locker".

In my opinion entertainment awards barely scrape the definition of news let alone trump the many real stories available today. BBC Radio 2 ran the story at the very end of its bulletin.

Useful? [Yes](#) (0) [No](#) (0)
[Report abuse](#) [Share](#) ▼

me News Sport Weather TV Radio

[Watch](#) **ONE-MINUTE WORLD NEWS**

Page last updated at 21:59 GMT, Monday, 14 June 2010 22:59 UK

LATEST: Fourth Colombian hostage found alive after three fellow capt

DEEPWATER HORIZON RESPONSE

Obama calls for clean energy push

US President Obama asks supporters to back a campaign for clean energy, as he visits areas affected by the BP oil spill.

- Mapping the oil slick
- In graphics: Tackling the leak
- Q&A: Scale of BP oil spill

Google Sidewiki Annotation on *http://news.bbc.co.uk/* as of 2010-06-14



Memento
SWIB 2011, Hamburg, Germany, November 2011

98



Web-Centric Annotation: No Persistence

Google Sidewiki

Robert Sanderson
[Edit your name](#)

Write an entry about: The entire page

To write an entry about a specific part of the page, highlight those words on the page.


This page has no Annotations :(

Helpful information about "BBC NEWS | News Front Page".

Page last updated at 07:04 GMT, Monday, 8 March 2010

LATEST: FBI releases some of its secret files on the late Senator Edward

Oscar joy for The Hurt Locker



Archived page from:

<http://www.dracos.co.uk/work/bbc-news-archive/2010/03/08/07.05.html>



Memento
SWIB 2011, Hamburg, Germany, November 2011

99



Web-Centric Annotation: Desired Persistence

[Paul Murray](#) - Mar 8, 2010

Lead story this morning.

With a magnitude six earthquake, attacks around the world killing hundreds and a domestic civil servants strike, I find it very odd that the BBC website should have chosen to lead with "Oscar triumph for The Hurt Locker".


In my opinion entertainment awards barely scrape the definition of news let alone trump the many real stories available today. BBC Radio 2 ran the story at the very end of its bulletin.

Useful? [Yes](#) (0) [No](#) (0)
[Report abuse](#) [Share](#) ▼

Page last updated at 07:04 GMT, Monday, 8 March 2010

LATEST: FBI releases some of its secret files on the late Senator

Oscar joy for The Hurt Locker





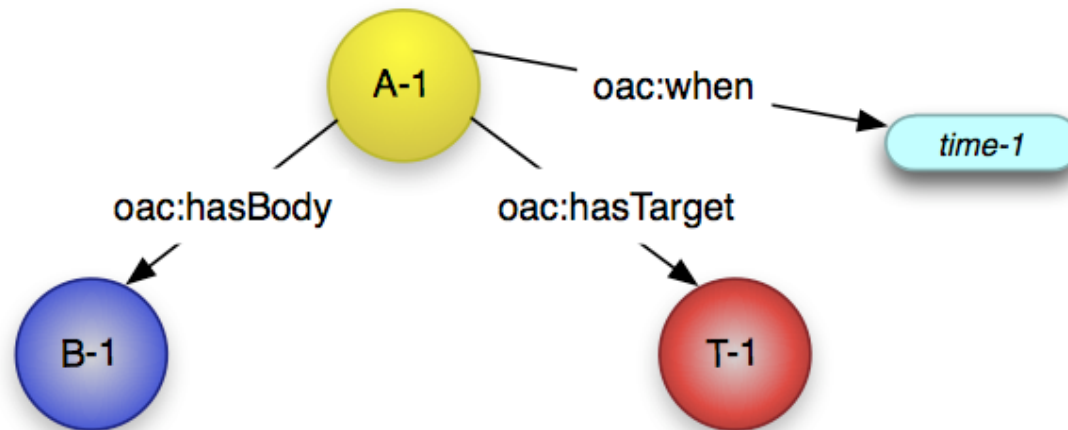
Open Annotation: Dealing with Web Time

- As regular Web resources, Body and Target of an Annotation have representations that can change over time.
- Body and Target can change independently of each other.
- If an Annotation involves resources as they existed at a particular point in time, this needs to be recorded.
- The OAC model provides hooks for doing so:
 - Timeless Annotations;
 - **Uniform Time Annotations;**
 - Varied Time Annotations.



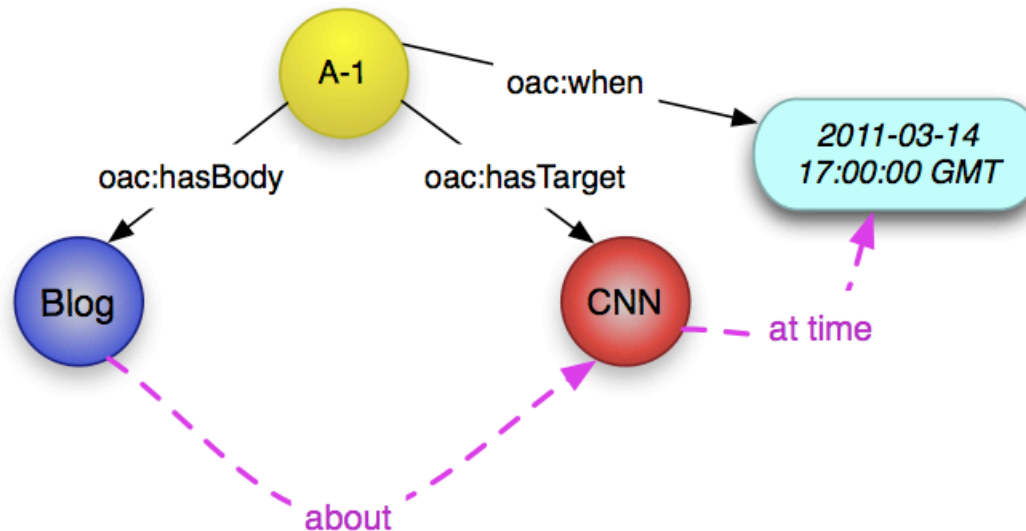
Open Annotation: Uniform Time Annotations

- The Annotation is not always applicable, but pertains to the state of the Body and Target at a specific moment in time.
- Add `oac:when` property to the Annotation.

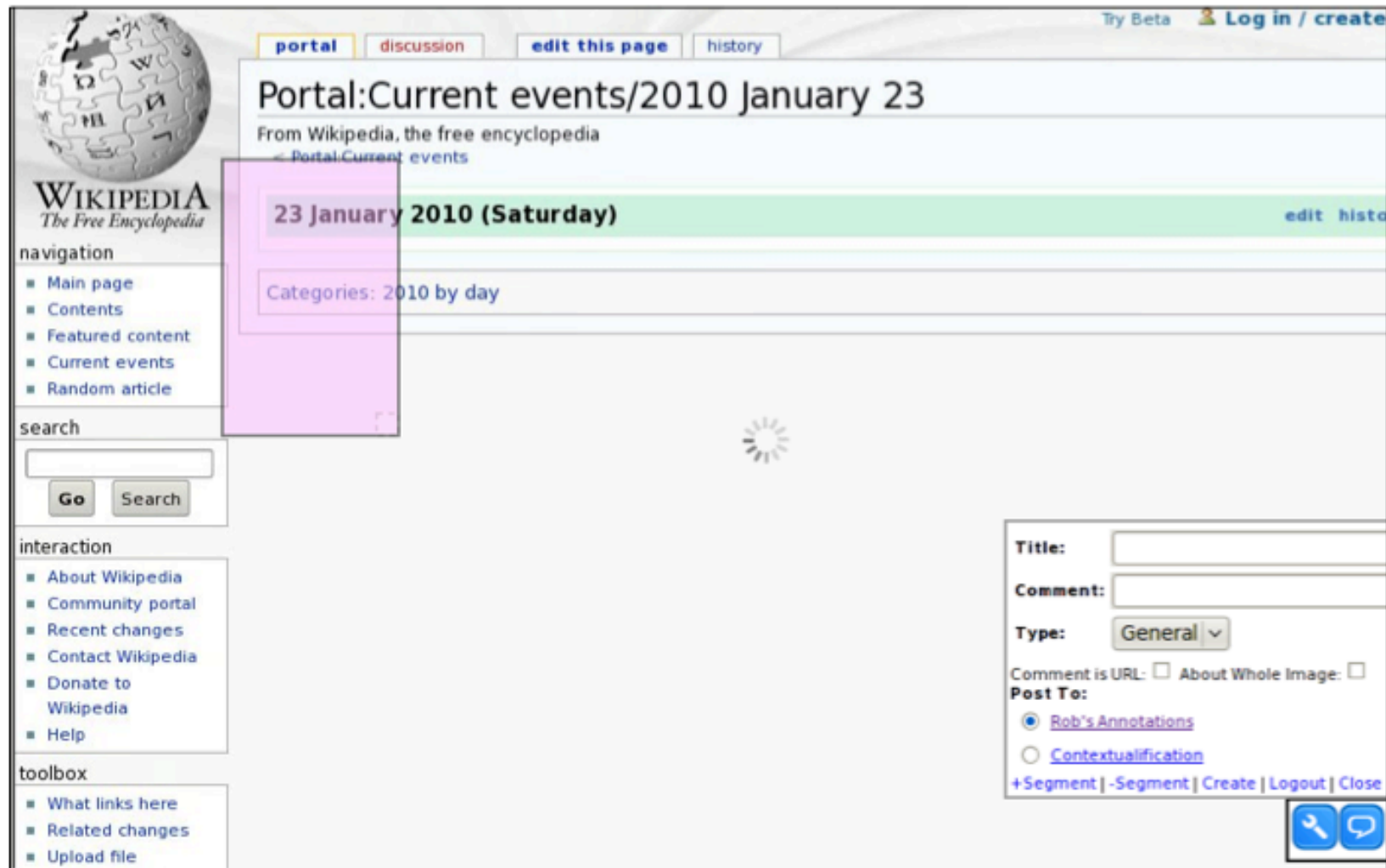


Memento + Open Annotation: Persistent Annotations

- In order to reconstruct the Annotation as intended: Use Memento to obtain an archived representation of B and T as they existed at the `oac:when` datetime.



Create an Annotation



Reconstruct the Annotation without Memento

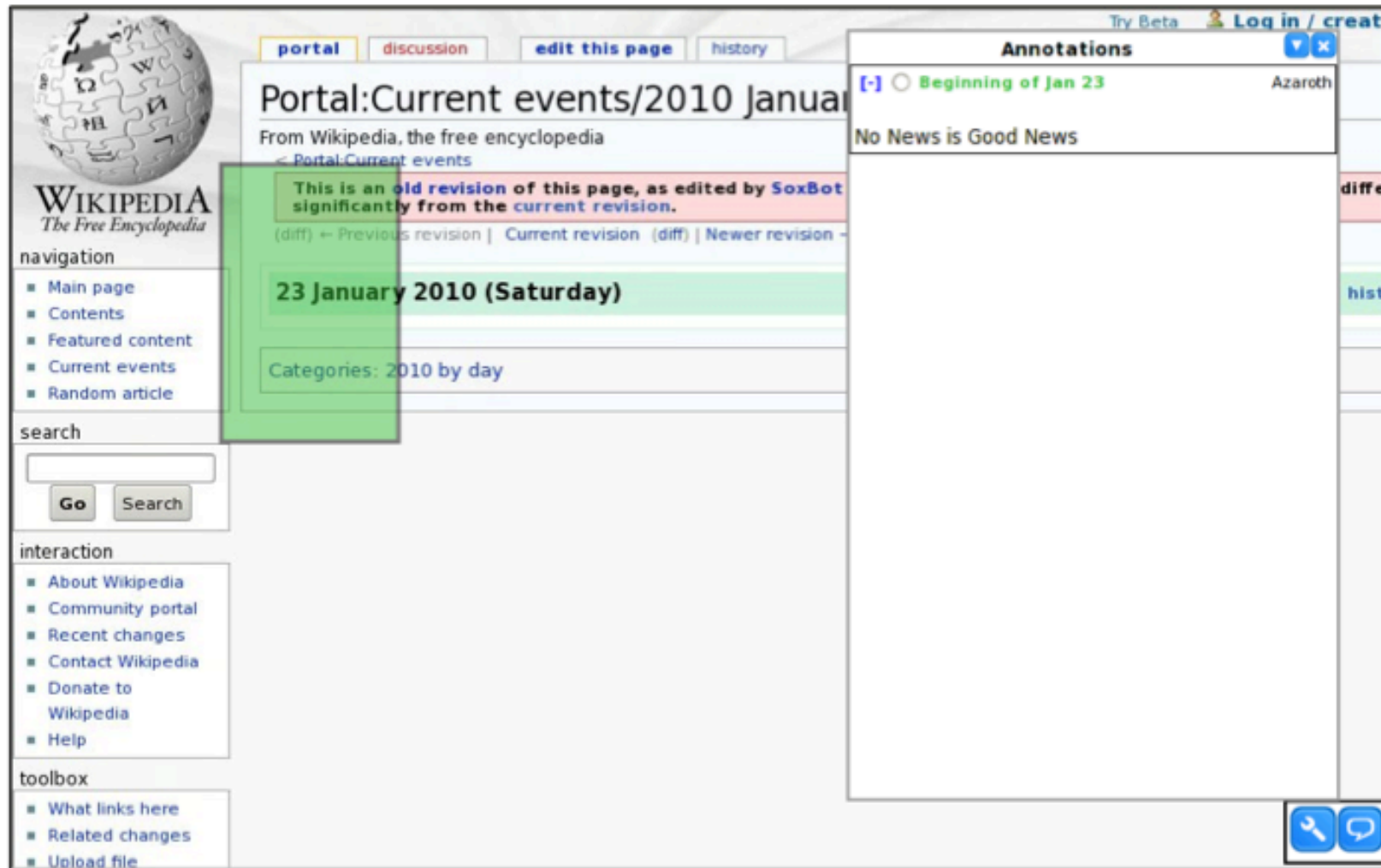


Memento
SWIB 2011, Hamburg, Germany, November 2011

105



Reconstruct the Annotation with Memento



paper at <http://arxiv.org/abs/1003.2643>



Memento
SWIB 2011, Hamburg, Germany, November 2011

106



The Memento Framework:

The Increasing Value of a URI



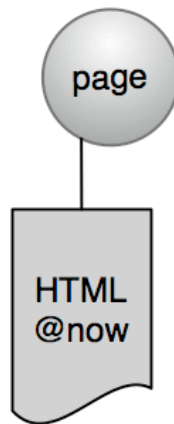
Memento
SWIB 2011, Hamburg, Germany, November 2011

107



URI as Access Point to a Page

<http://weather.example.com/oxaca>

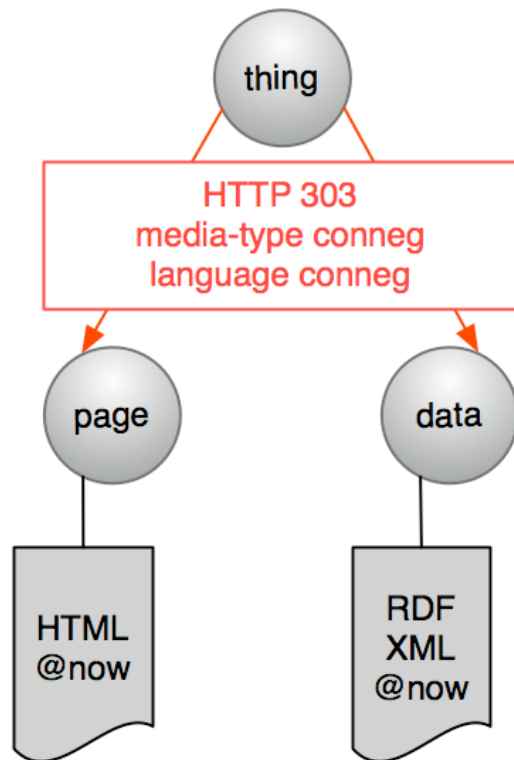


Memento
SWIB 2011, Hamburg, Germany, November 2011

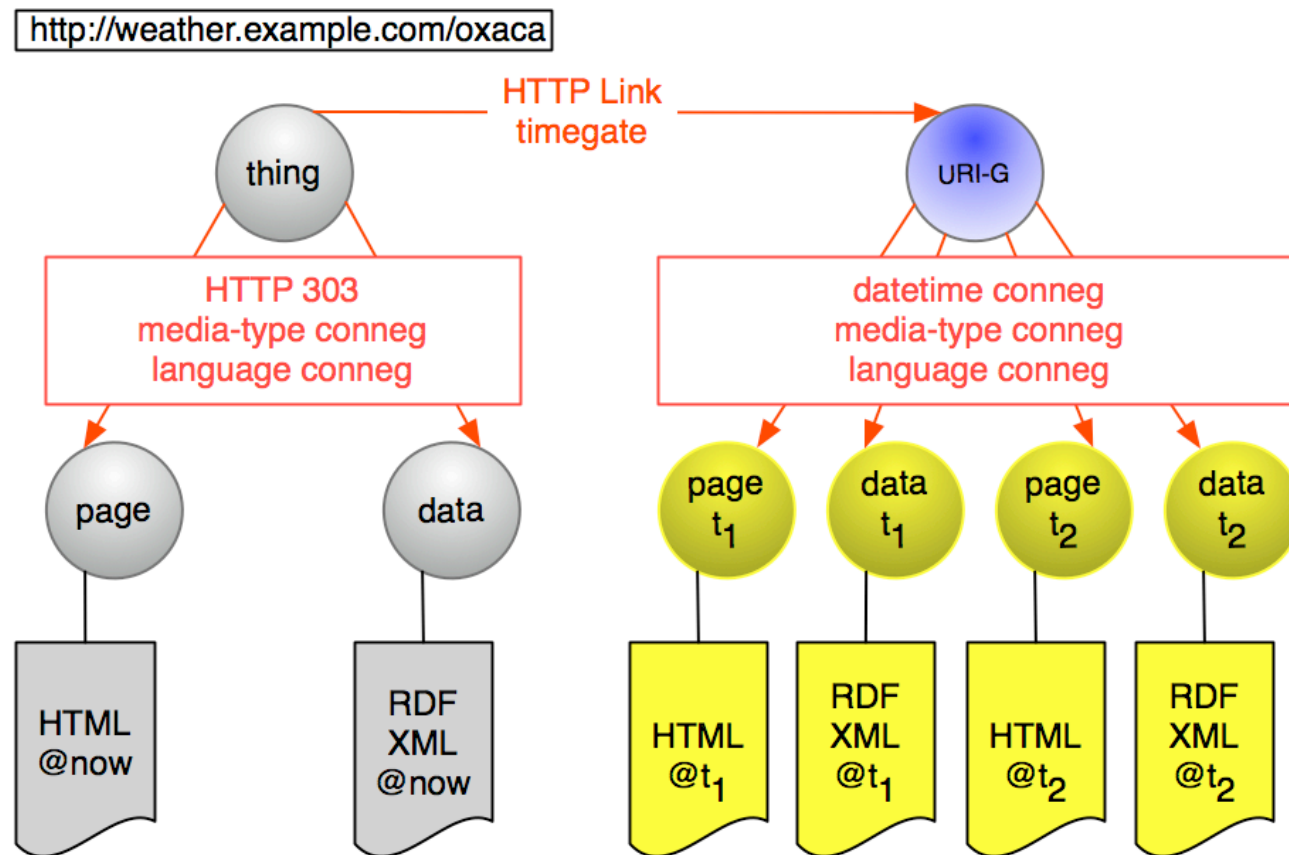


URI as Access Point to Page and Data

<http://weather.example.com/oxaca>



URI as Access Point to Current and Past Pages and Data



The Memento Framework:

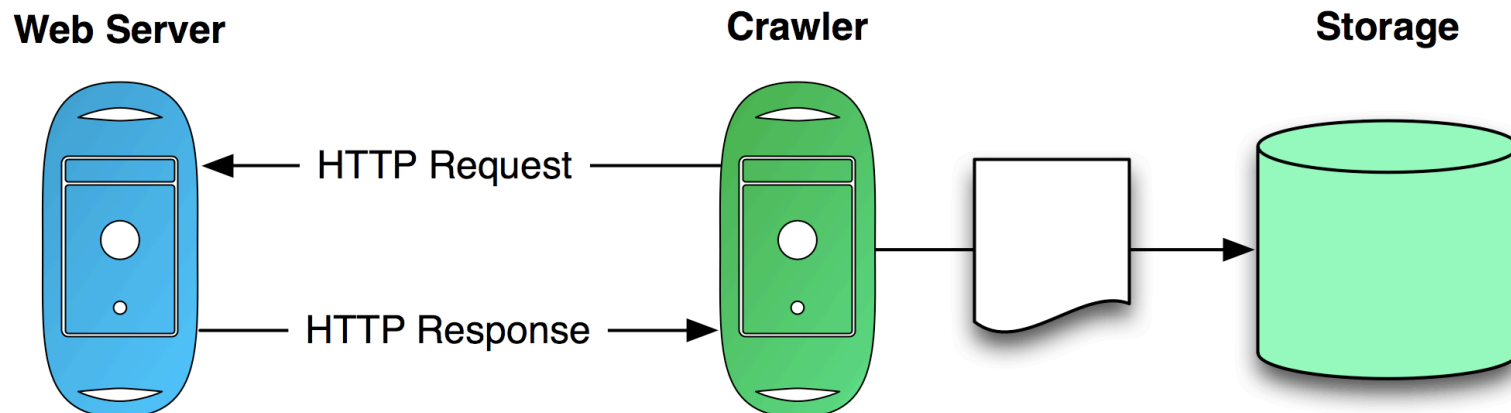
Transactional Archives



Memento
SWIB 2011, Hamburg, Germany, November 2011 ¹¹¹



Crawl-Based Web Archives



Observations

For example: Heritrix crawler for Internet Archive

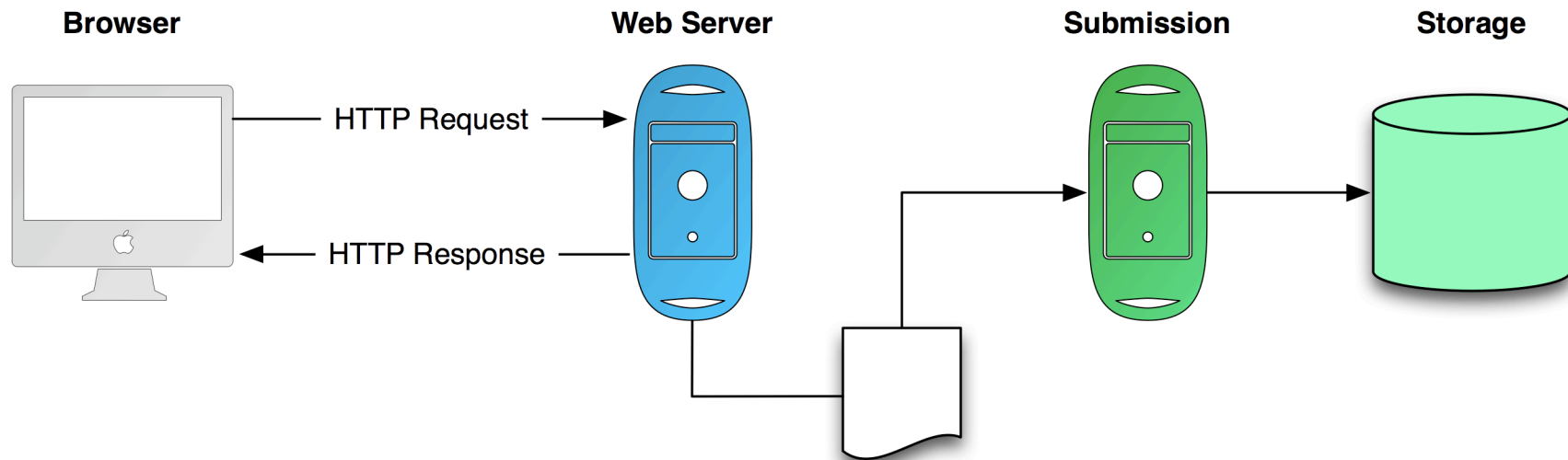


Crawl-Based Web Archives

- Collect discreet observations of resources, not their entire evolution.
- Can be rejected (`robots.txt`, by user-agent, by host IP)
- Can be deceived (cloaking, by geo-location, by user-agent).
- Coverage of particular Web server dependent on crawl-strategy.



Server-Side Transactional Web Archives



Change History

For example: TTApatch, PageVault, Vignette Web Capture



Server-Side Transactional Web Archives

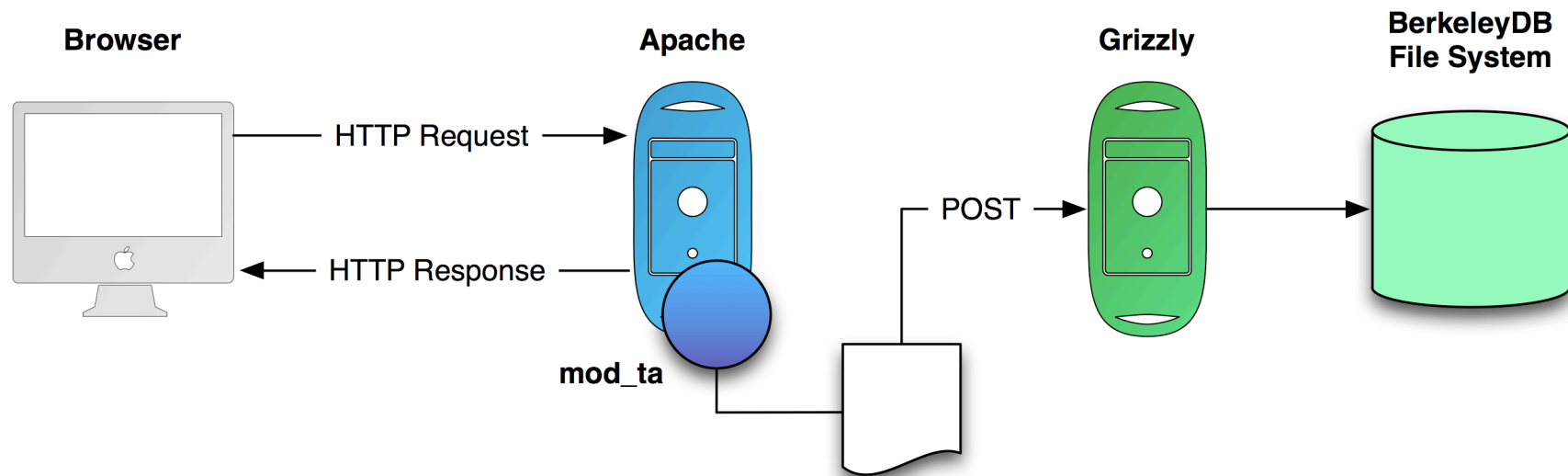
- Collect all representations served by to-be-archived server.
- To-be-archived server needs to cooperate.
 - Incentives e.g. institutional memory, official record of Web presence.
- Archival coverage restricted by to-be-archived server, does not include external servers (e.g. embedded resources).
- To be archived server can submit falsified information.
- Archival collection management: what to keep, what not (e.g. significant changes, deduplication, ...).



Development of Transactional Web Archive Software

Capture:

- Apache connection filter module captures URI, headers, body
- POSTs in real-time to transactional archive



Access:

- Online, real time access via Memento TimeGates
- Batch Export via WARC files for long term preservation
- Check out <http://theresourcedepot.org>



References

- Van de Sompel, H., Nelson, M.L., Sanderson, R., Balakireva, L., Ainsworth, S., Shankar, H. (2009) Memento. <http://arxiv.org/abs/0911.1112>
- Van de Sompel, H., Sanderson, R., Nelson, M.L., Balakireva, L., Ainsworth, S., Shankar, H. (2010) An HTTP-Based Versioning Mechanism for Linked Data. Proceedings of the 3rd Workshop on Linked Data on the Web. <http://arxiv.org/abs/1003.3661>
- Sanderson, R., and Van de Sompel, H. (2010) Making Web Annotations Persistent over Time. Proceedings of the 10th ACM/IEEE-CS Joint Conference on Digital libraries. <http://arxiv.org/abs/1003.2643>
- Sanderson, R., Phillips, M. and Van de Sompel, H. (2011) Analyzing the persistence of referenced Web resources with Memento. OR 2011. <http://arxiv.org/abs/1105.3459>
- Van de Sompel, H., Nelson, M.L., Sanderson, R. (2011) HTTP based framework for time-based access to resource states - Memento <http://www.mementoweb.org/guide/rfc/ID/>



Memento wants to make navigating the Web's Past Easy



<http://mementoweb.org/>
<http://groups.google.com/group/memento-dev>



Memento
SWIB 2011, Hamburg, Germany, November 2011

118

