



Web Architecture

Frank McCown

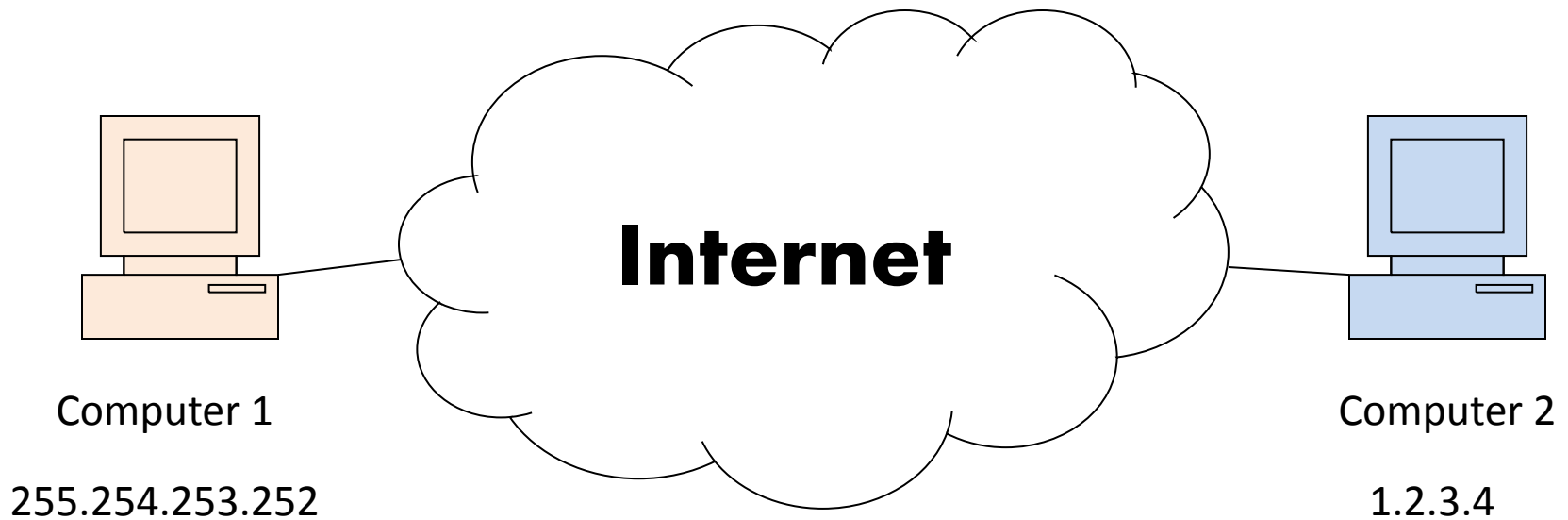
Introduction to Web Science

Harding University

Spring 2011

“The **Internet** is a global system of interconnected computer networks that use the standard Internet Protocol Suite (TCP/IP) to serve billions of users worldwide.”

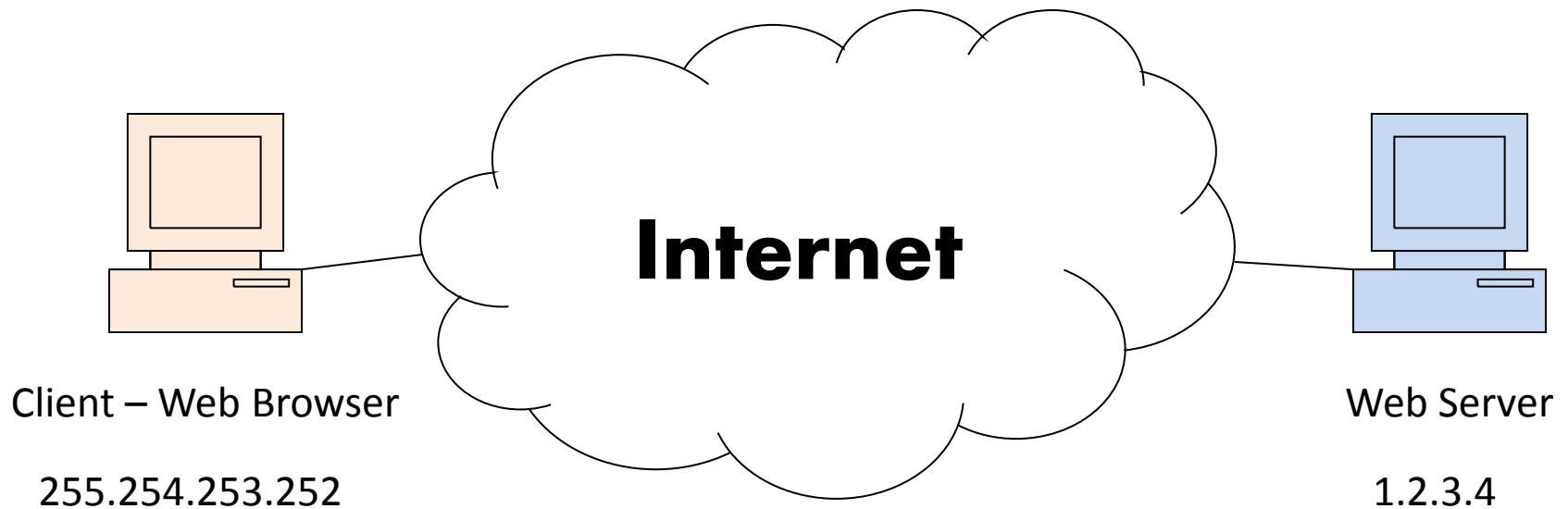
<http://en.wikipedia.org/wiki/Internet>



Internet Protocol Suite

- **Internet Protocol (IP)**: directs packets to a specific computer using an IP address
- **Transmission Control Protocol (TCP)**: directs packets to a specific application on a computer using a port number.
 - Common port numbers:
 - 22 – ssh
 - 23 – telnet
 - 25 – email
 - 80 – Web

Overview of the Web



World Wide Web: The system of interlinked hypertext documents accessed over the Internet using the HTTP protocol.

Hypertext Transfer Protocol (HTTP)

HTTP is the set of rules that govern communication between web browsers and web servers.

Client Request

```
GET /comp/ HTTP/1.1  
Host: www.harding.edu
```

Example request for

<http://www.harding.edu/comp/>

Server Response

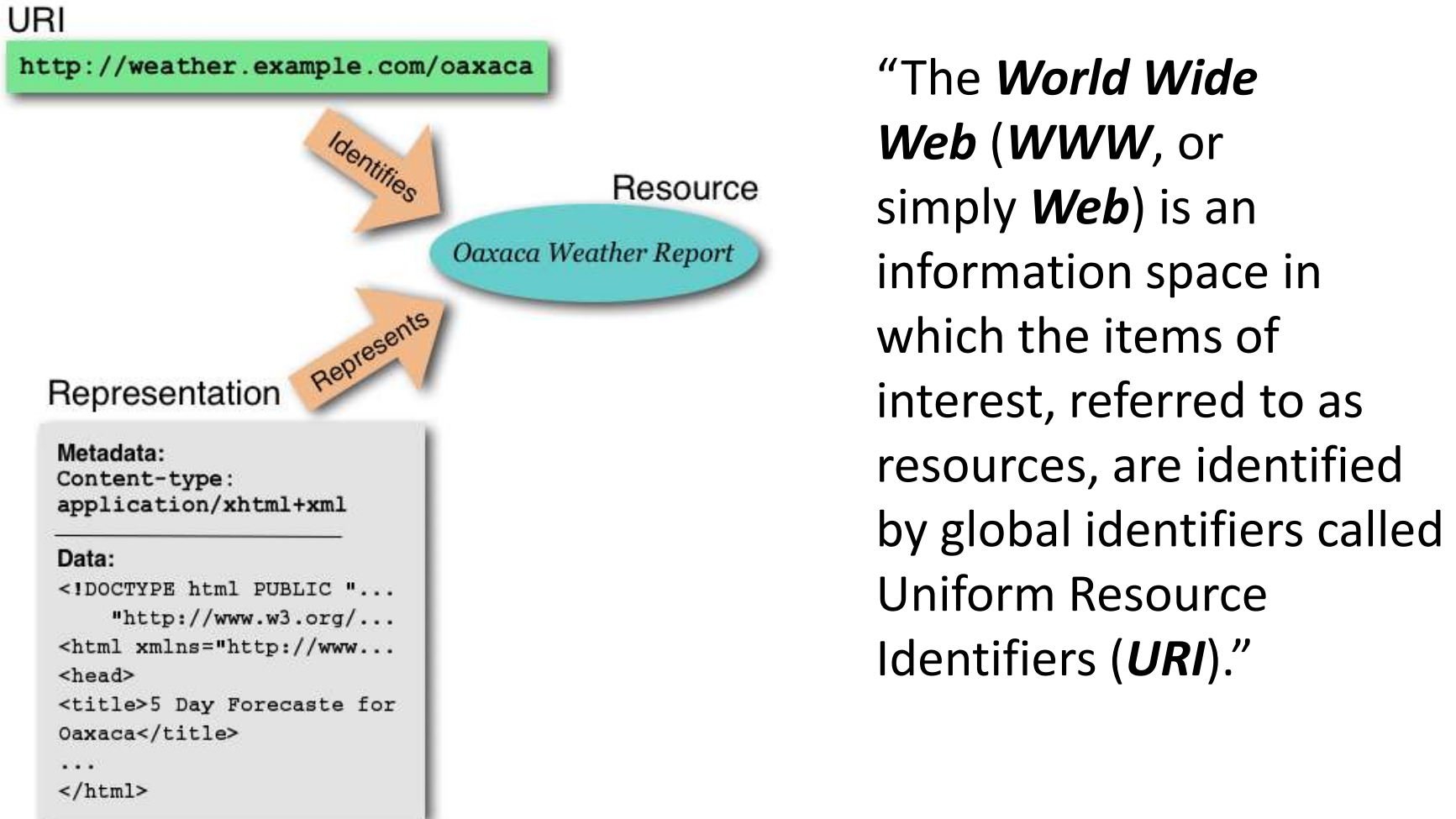
```
HTTP/1.1 200 OK  
Content-Length: 6018  
Content-Type: text/html  
Content-Location: http://www.harding.edu/comp/  
Last-Modified: Mon, 05 Jul 2010 18:49:40 GMT  
Server: Microsoft-IIS/6.0
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN"  
"http://www.w3.org/TR/html4/strict.dtd"> <html> <head>  
<title>Harding University - Computer Science</title>
```

More Formal Definitions

- HTTP defined by Request for Comments (RFCs) 1945, 2068, and 2616
- Other RFCs for defining URLs (1736, 1738), URIs (1630, 2396), etc.
- Web architecture defined in W3C's *The Architecture of the World Wide Web, Volume One*
 - <http://www.w3.org/TR/webarch/>

Web Definition



“The *World Wide Web* (*WWW*, or simply *Web*) is an information space in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers (*URI*).”

URIs, Resources, and Representations

URI

`http://weather.example.com/oaxaca`

Identifies

Resource

Oaxaca Weather Report

Represents

Representation

Metadata:

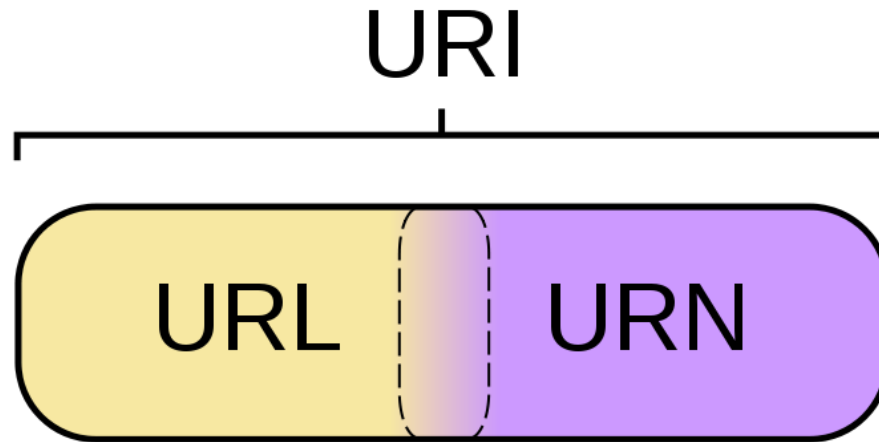
Content-type:
application/xhtml+xml

Data:

```
<!DOCTYPE html PUBLIC "...  
    "http://www.w3.org/...  
<html xmlns="http://www...  
<head>  
<title>5 Day Forecaste for  
Oaxaca</title>  
...  
</html>
```

- URIs *identify* Resources
- Representations *represent* Resources
- When URIs are dereferenced, they return representations (not resources)
- Different representations may be returned for the same URI (e.g., English vs. French version)

What's a URx?



URI - String of chars used to identify a name or resource on the Internet

URL - Where to find a resource

URN - Name of a resource

Content Negotiation

- Content negotiation (RFC 2616 sec 12) is used in HTTP to allow servers to send different representations of the same resource at the same URI
- The user agent tells the server which encoding, language, media type, etc. it prefers, and the server responds with the “best” representation
- Example HTTP request headers:

```
Accept-Language: fr; q=1.0, en; q=0.5  
Accept: text/html; q=1.0, text/*; q=0.8,  
       image/gif; q=0.6, image/jpeg; q=0.6, image/*;  
       q=0.5, */*; q=0.1
```

This agent prefers French over English, HTML over other document types, and GIF and JPEG over other image formats

Memento – Date/Time Negotiation

- Memento is a new protocol which uses HTTP content negotiation to retrieve older versions (Mementos) of web resources
- Agent request URI with Accept-Datetime set to desired date/time
- Server responds with a link to a TimeGate which knows the Mementos available for the URI
- Agent makes request to TimeGate and receives response with the URL to the Memento
- Learn more: <http://mementoweb.org/>

