

COMP6218 Web Architecture

- Introduction and Discussion
 - What is the Web?
 - What makes it so “Webby”?
 - What was new about it that we didn’t have before?
 - What is the USP of the Web?

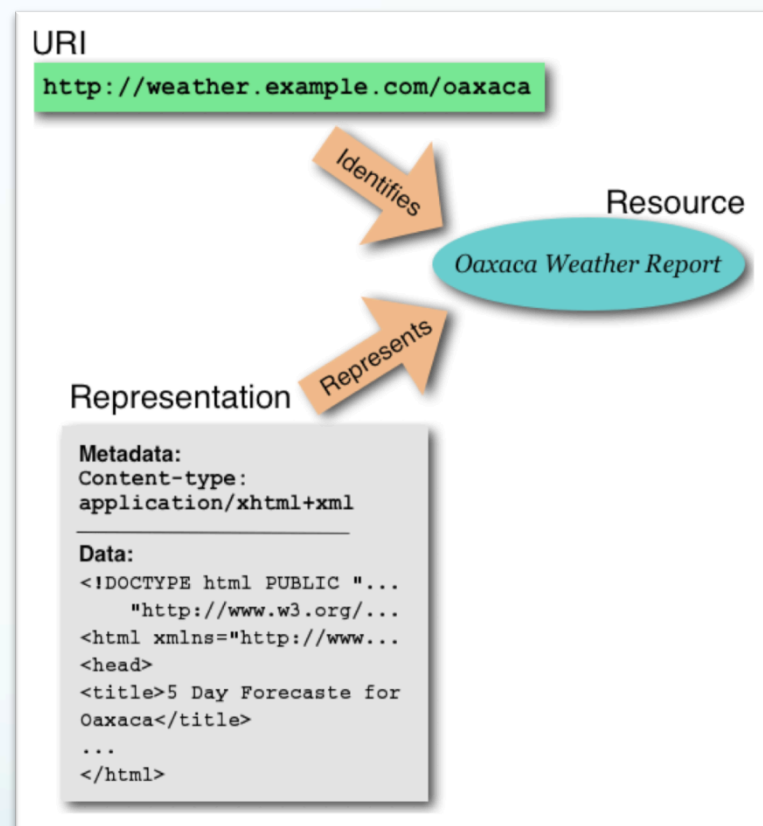
How Does the Web Work?

- This man is reading the Guardian newspaper on the Web.
- What technology underpins his activity?
- EXERCISE: Brainstorm all the programs, protocols, standards, data formats and TLAs you can think of that contribute to the Web as you use it.

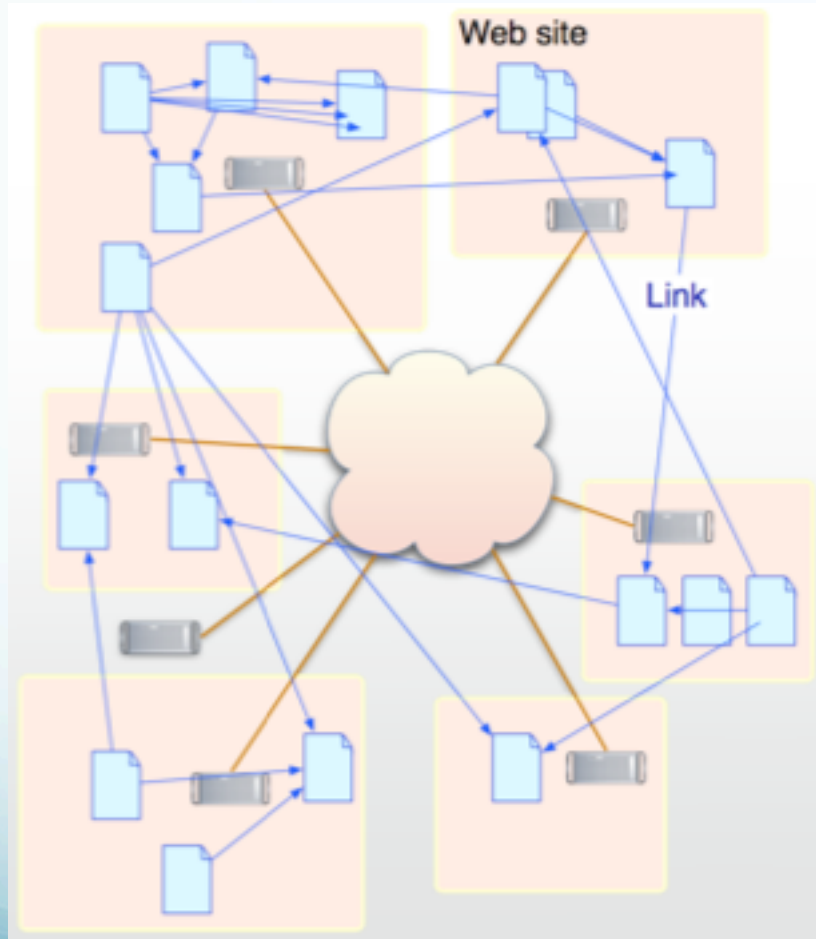


Web Architecture

- Resources are identified by URIs
- Resources have different representations in different formats (e.g. HTML, text, PDF)
- Resources can be obtained by asking the Internet for them (using HTTP)
- **Key components of the Web Architecture:**
 - Identification
 - Interaction
 - Formats

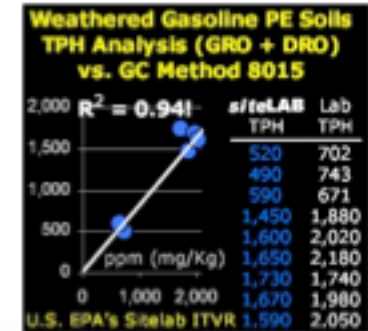
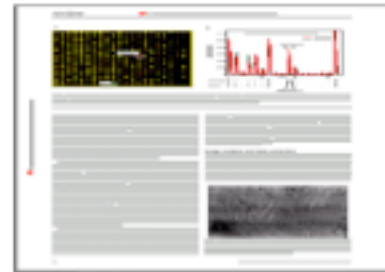


Web Principles: Web of Documents and Data

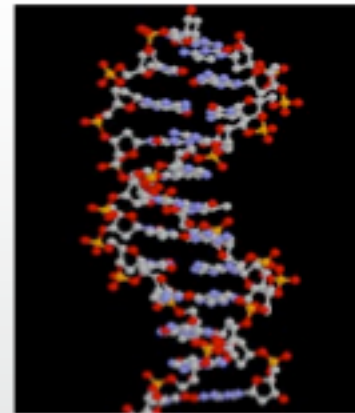


URIs identify *any* resource

- Publications
- Multimedia
- Web data set (XHTML)
- Databases
- Scientific structures
- Workflows
- People



Web data set (XHTML)



Web Principles

- All entities of interest, such as information resources, real-world objects, and vocabulary terms should be identified by URI references
- URI references should be *dereferenceable*, meaning that an application can look up a URI over the HTTP protocol and retrieve data about the identified resource (a representation).
- Data should be provided using a standard format (HTML, XML, RDF etc)
- Data should be interlinked with other data

5 Stars of Linked Data (2010)

★ Available on the web (whatever format), but with an open licence

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

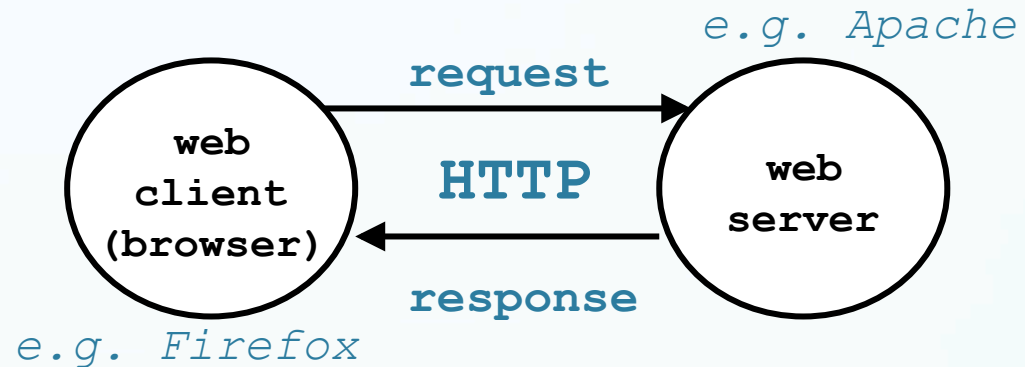
★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)

★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context

The Web Experience

- A user clicks on a link in a browser.
- The browser communicates with a web server using HTTP
- The server sends an HTML document back
- The browser displays the document
- The user clicks on another link and activates another URL



Pre Web: File Transfer

- A user typed a host address into a client.
- The client communicated with a file server using File Transfer Protocol (FTP)
- The user typed commands into the client
 - to navigate to the right directory
 - to GET the right file from a DIR listing
 - to specify BINARY or ASCII transfers to make sure that line endings were treated correctly.
- The server sent a PostScript or text document back
- The client **stored the document on the hard disk**
- The user **printed** the document

Pre Web: FTP

FTP commands

PostScript data

```
Remote system type is Windows_NT.
ftp> ls
200 PORT command successful.
150 Opening ASCII mode data connection for /bin/ls.
11-22-05 03:40PM <DIR> MSC
10-20-06 11:55AM <DIR> Rob
226 Transfer complete.
ftp> cd MSC
250 CWD command successful.
ftp> ls
200 PORT command successful.
150 Opening ASCII mode data connection for /bin/ls.
11-22-05 03:40PM <DIR> Aaa
11-22-05 03:40PM <DIR> Alice
09-06-00 02:31PM 13478006 arc explorer.exe
01-31-00 12:00AM 1441792 archdata.mdb
11-22-05 03:40PM <DIR> blaster
11-22-05 03:40PM <DIR> courses
07-04-02 04:17PM 12612 Curtin-Aleph500.enz
11-22-05 03:40PM <DIR> Distance_education
02-27-02 12:52PM 19456 Doc1.doc
11-22-05 03:40PM <DIR> documents
07-30-02 10:50AM 95744 docxplorer.doc
07-31-03 10:24AM 1502 error.txt
11-22-05 03:40PM <DIR> Examples
02-27-02 10:18AM 105984 ftp-http-protocols.doc
08-04-03 08:01AM 15646208 infotrove_nt_2_15.exe
07-30-02 08:26AM 1018090 Infotrove_user_manual.pdf
11-22-05 03:40PM <DIR> journalism
03-16-01 09:58AM 12835 Library Catalog (CDL).enz
11-22-05 03:40PM <DIR> mc1111
11-22-05 03:40PM 3062925 n10e30.exe
04-13-02 12:10PM 81400 PhpProposal.doc
07-05-01 11:04AM 1956 readme.txt
03-11-02 09:02AM 5656 sample_data.sav
11-22-05 03:40PM <DIR> SOHY
11-22-05 03:40PM <DIR> TMAGIC
05-14-99 12:08PM 3743302 tmagic.zip
02-18-02 10:00AM 13680 topic8-supplement.html
11-22-05 03:40PM <DIR> wethu
226 Transfer complete.
ftp> get readme.txt
local: readme.txt remote: readme.txt
200 PORT command successful.
150 Opening ASCII mode data connection for readme.txt(1956 bytes).
WARNING! 60 bare linefeeds received in ASCII mode
File may not have transferred correctly.
226 Transfer complete.
1956 bytes received in 0.01 secs (137.1 kB/s)
ftp> |
```

- Pre web interaction was characterised by **DOWNLOADING** instead of **BROWSING**.

```
%%BeginSetup
%%Feature: *Resolution 600dpi
TeXDict begin

%%EndSetup
%%Page: 1 1^M
/Times-Roman findfont
12 scalefont
setfont
newpath
131 562 moveto
(Hello, world!) show
131 542 moveto
(Strickly speaking, \
wavelets are topic of pure mathematics!!!!) show
/Helvetica findfont
16 scalefont
setfont
131 522 moveto
(Does the line above look better than the one below?) show
1 0 bop 1339 271 a FH(M)-15 b(A)g(VELETS)44 b(F)l(OR)g(KIDS)1470
391 y FG(A)37 b(T)-9 b(utorial)36 b(In)(tro)(duction)1986
812 y FF(Bj)992 932 y(Brani)i(Uid)n(ak)n(o)n(vic)201
b FE(and)162 b FF(Peter)37 b(Mueller)1711 1052 y FD(Duke)e(University)
"kidsE.ps" [unix] 3414L, 231145C written 3366,1 99%
```



User types commands directly to server. Document transferred between hard disks. User prints the file to read its contents.



Consolidating Learning



Mary, Mary, quite contrary,
How does your garden grow?

With silver bells and cockle
shells

And pretty maids all in a row

The Web Architecture



All together now!

TimBL, TimBL, very nimble,
How does your linked Web
grow?

With URLs and HTMLs

And GET & POSTs all in a row

Something to think about...

- If we agree that *everything we can imagine in the whole wide world* should have an address and a way to look it up and get hold of it...
- ...haven't we just destroyed privacy? secrecy?