



University of
Southampton

COMP3227 Web Architecture and Hypertext Technologies

Dr Heather Packer – hp3@ecs.soton.ac.uk

What's this module about?

- What is the Web?
- How is the Web made?
- What came before the Web?
- Where is the Web going?

What's this module *not* about?

- How to write HTML (although we will look at its evolution and capabilities)
- How to write CSS (although we will look at its capabilities)
- How to set up a web server
- How to write applications in PHP/ASP.NET/Ruby on Rails/Django/node.js

Other ECS modules cover some of this, most notably COMP6205 Web Development
Some material is covered in more depth in COMP6215 Semantic Web Technologies

Module structure

Three lectures per week:

- Monday 2pm in 85/2209
- Wednesday 10am in 7/3023
- Friday 4pm in 85/2209

One lab per week (starting in Week 3):

- Thursday 10am in 59/ECS Computing Lab

Links to all module resources will be at
<https://secure.ecs.soton.ac.uk/module/COMP3227/>

Laboratories

Starting Week 3 (w/c 17 Oct 2021)

The first four labs have formative exercises to prepare you for the coursework:

- HTTP
- HTML
- CSS
- REST

Subsequent labs will run as coursework clinics

Assessment

Examination: 50% (120 minutes, 3 questions from 4)

REST architecture coursework: 50%

- Specification published in week 1
- Submission due week 9
- Feedback due week 15

Week-by-week topics

- Week 1: Hypertext and the Architecture of the Web (hsp/rcg)
- Week 2: Web Protocols (hsp/rcg)
- Week 3: Web Formats (hsp/rcg)
- Week 4: Styling the Web, and Advanced Protocols (hsp/rcg)
- Week 5: RESTful Web Services (hsp/rcg)
- Week 6: History of Hypertext (nmg)
- Week 7: Open Hypermedia (hsp/rcg)
- Week 8: Web Graph and Search Engines (hsp/rcg)
- Week 9: Caching, Content Delivery and Web Advertising (hsp/rcg)
- Week 10: Linked Data, Open Data and Open Access (hsp/rcg)
- Week 11: Intellectual Property and Net Neutrality (rcg)
- Week 15: Review (hsp/rcg)

A note on assessable content

Not everything we cover on this module can be sensibly assessed, so we'll indicate topics of which you only need a high-level understanding as follows.

If a topic concerns technologies which are currently undergoing standardisation (and so which are subject to change), we'll indicate it like this:

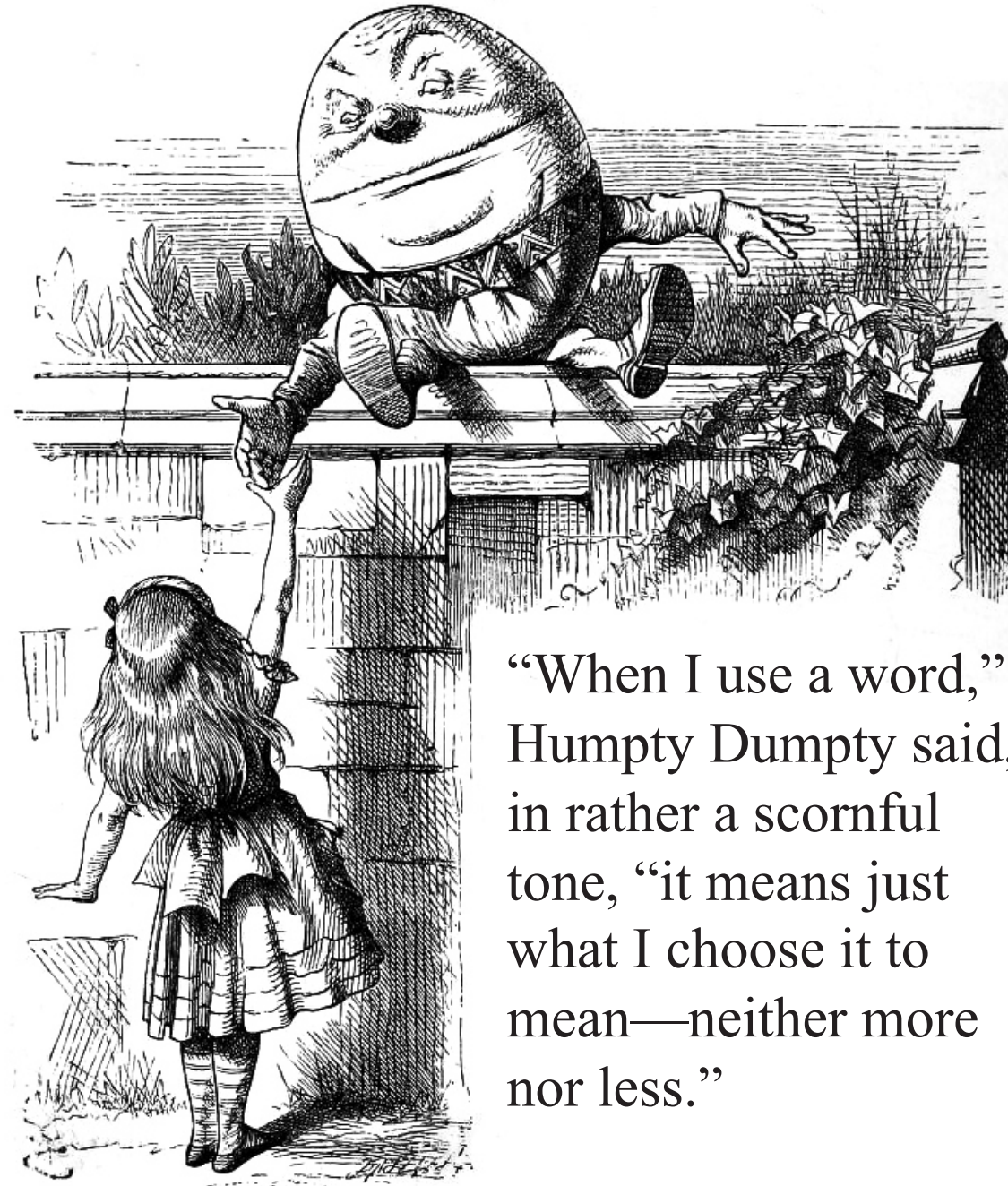


If a topic goes into large amounts of detail that's primarily there only for illustrative purposes, we'll indicate it like this:



The World Wide Web

What is the World Wide Web?



“When I use a word,”
Humpty Dumpty said,
in rather a scornful
tone, “it means just
what I choose it to
mean—neither more
nor less.”

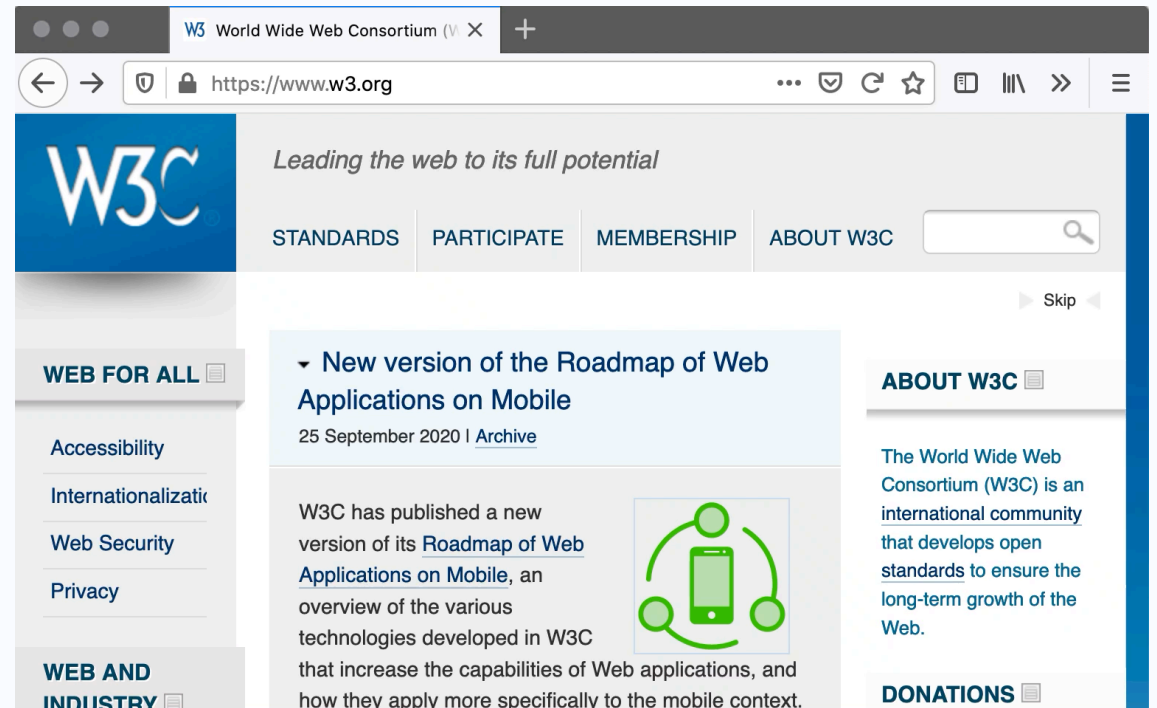
Before the Web

- A user typed a host address into a client program
- The client communicated with a file server using the File Transfer Protocol (FTP)
- The user typed commands into the client:
 - To navigate to the right directory
 - To specify whether the file being transferred was binary or ASCII
 - To get the right file
- The server sent a file back
- The client stored the file on the hard disk
- The user printed the file, or used a separate viewer



The Web experience

- A user clicks on a link in a browser
- (the browser talks to a web server)
- (the server sends a document back)
- The browser displays the document
- The user clicks on another link (etc)



Web evolution

The Web is for scientists (1991-1995)

The Web is for commerce (1996-2000)

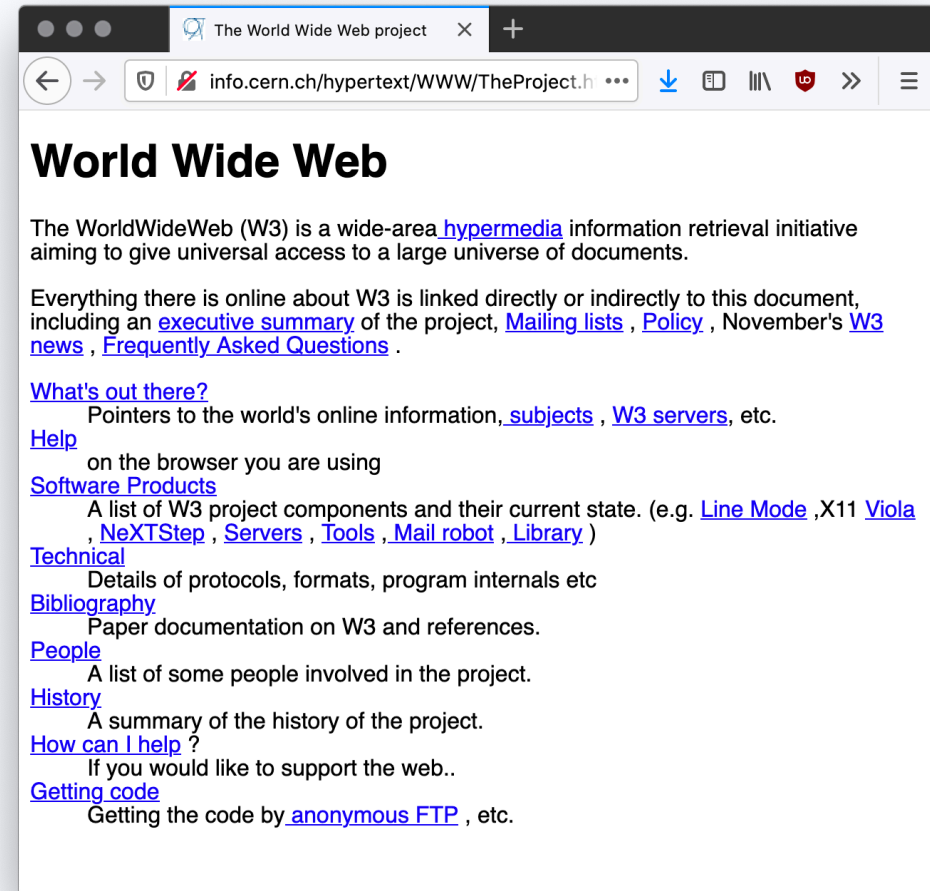
The Web is for users (2000-2005)

The Web as application platform (2005-)

Web evolution

The Web is for scientists (1991-1995)

- Berners Lee 1991
- High Energy Physics and Computer Scientists
- Library science community
- Document-centric
 - Web pages
 - PDF and PostScript files
- Limited interactivity
- Many browsers



Web evolution

The Web is for scientists (1991-1995)

The Web is for commerce (1996-2000)

- Mosaic browser developed at the National Center for Supercomputing Applications at the University of Illinois and Urbana Champaign
- Invention of Secure Sockets
- Netscape released JavaScript
- Microsoft released Cascading Style Sheets (CSS)
- The dot-com bubble



Web Evolution

The Web is for scientists (1991-1995)

The Web is for commerce (1996-2000)

The Web is for users (2000-2005)

The Web as application platform (2005-)

- Semantic Web
- W3C vs Browser manufacturers
- HTML5 Application platform
- Shift in emphasis from documents to applications

HTML



What is the World Wide Web?

A **distributed information system** that provides access to hypertext documents and other objects of interest



We have a general name for these objects of interest:

resources

What is a resource?

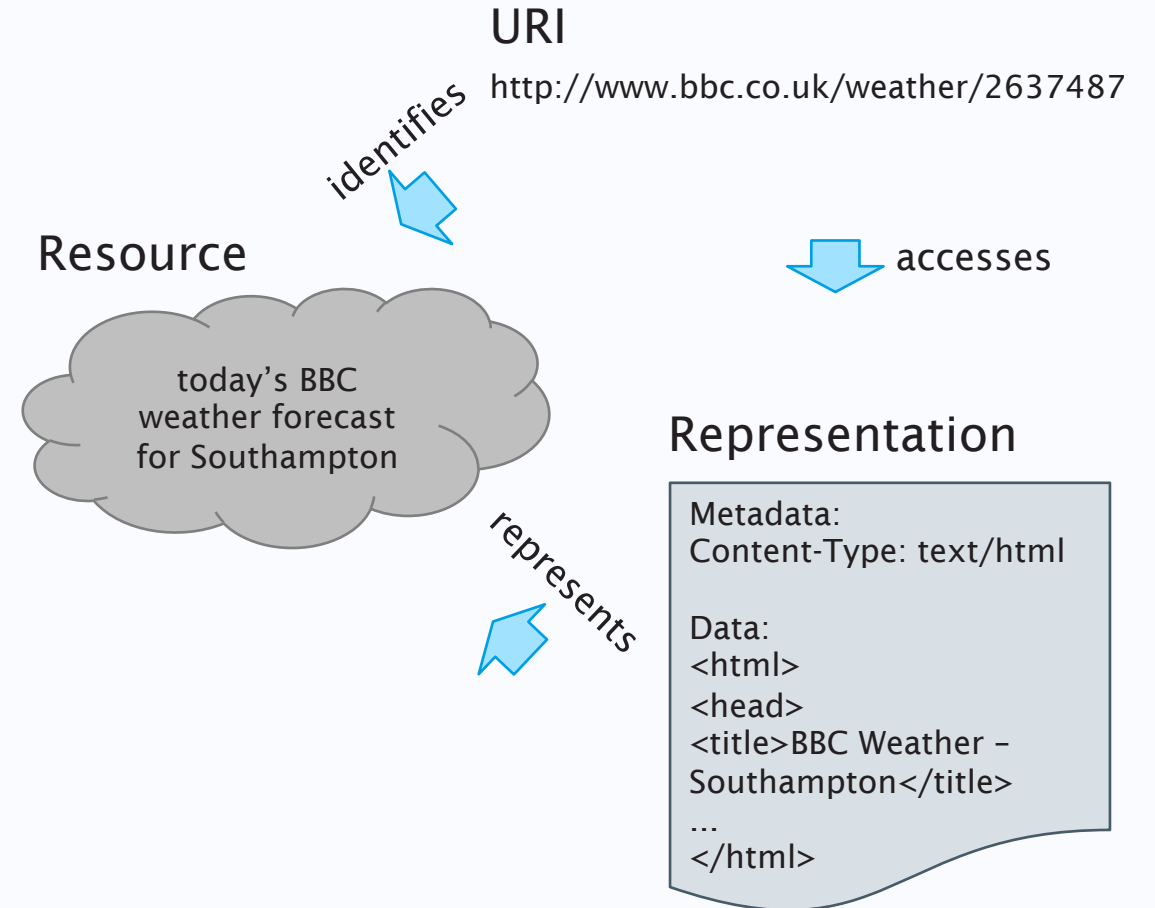
“Familiar examples [of resources] include an **electronic document**, an **image**, a **source of information** with a consistent purpose (e.g., ‘today’s weather report for Los Angeles’), a **service** (e.g., an HTTP-to-SMS gateway), and a **collection of other resources**. A resource is **not necessarily accessible via the Internet**; e.g., human beings, corporations, and bound books in a library can also be resources. Likewise, **abstract concepts** can be resources, such as the **operators** and **operands** of a **mathematical equation**, the types of a **relationship** (e.g., ‘parent’ or ‘employee’), or **numeric values** (e.g., zero, one, and infinity).”

Web Architecture

Resources are *identified* by URIs
(Uniform Resource Identifiers)

Resources have *representations* in
different formats (HTML, text, PDF)

Resources can be *interacted* with using
network protocols (HTTP)



Web Principles

- All entities of interest should be identified by URIs
- All URIs should be resolvable (i.e. you can use them to fetch something)
- When you resolve a URI, you get some data about the identified resource
- Data should be provided using standard formats
- Data should be linked with other data

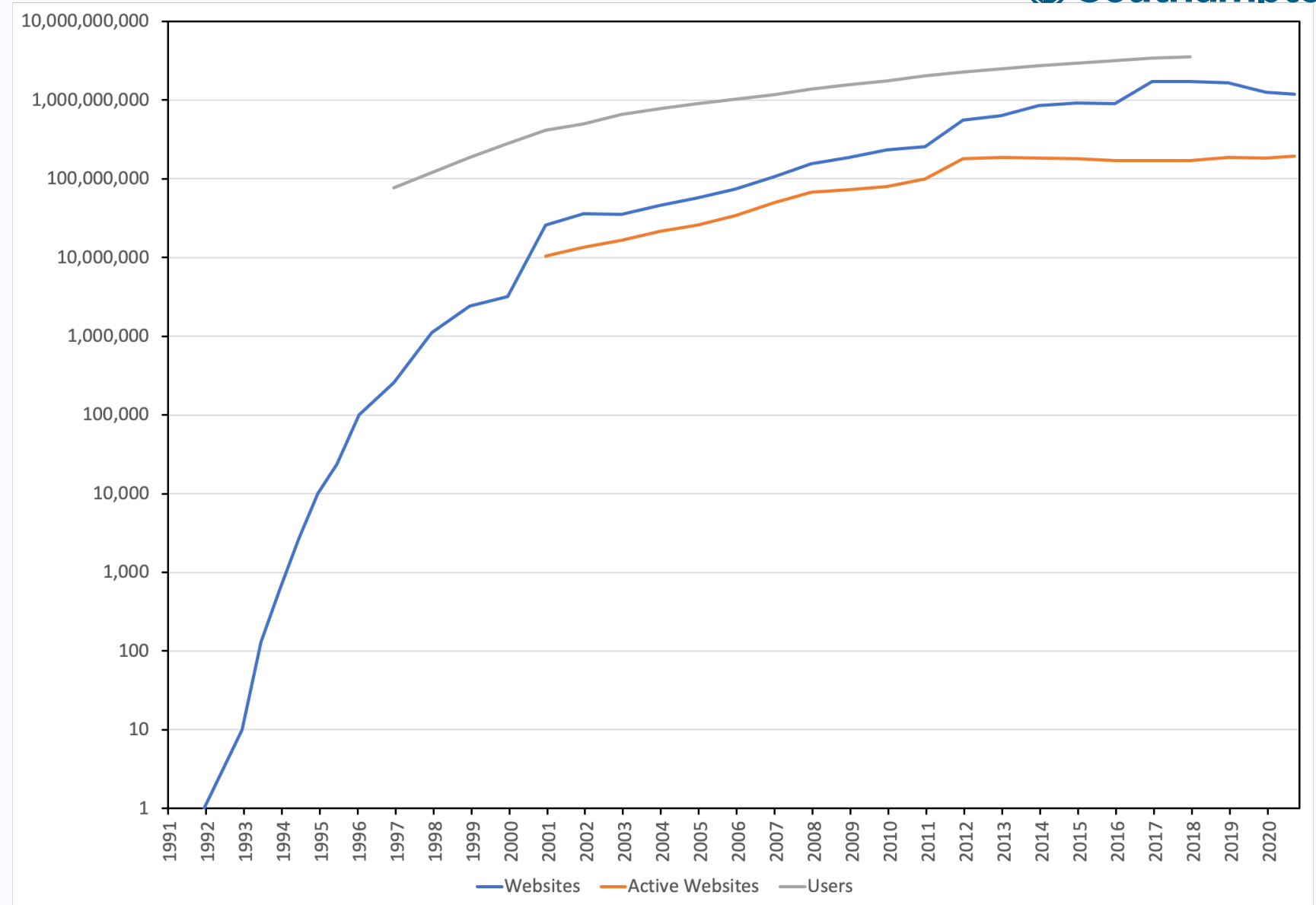
5 Stars of Linked Data (2010) for Organisations

- ★ Available on the Web (in whatever format) under an open licence
- ★★ As above, but as machine-readable structured data (e.g. Excel instead of an image of a table)
- ★★★ As above, but in a non-proprietary format (e.g. CSV instead of Excel)
- ★★★★ As above, but using W3C standards (RDF, SPARQL) to identify things, so that others can point at your data
- ★★★★★ As above, but linked to other people's data to provide context

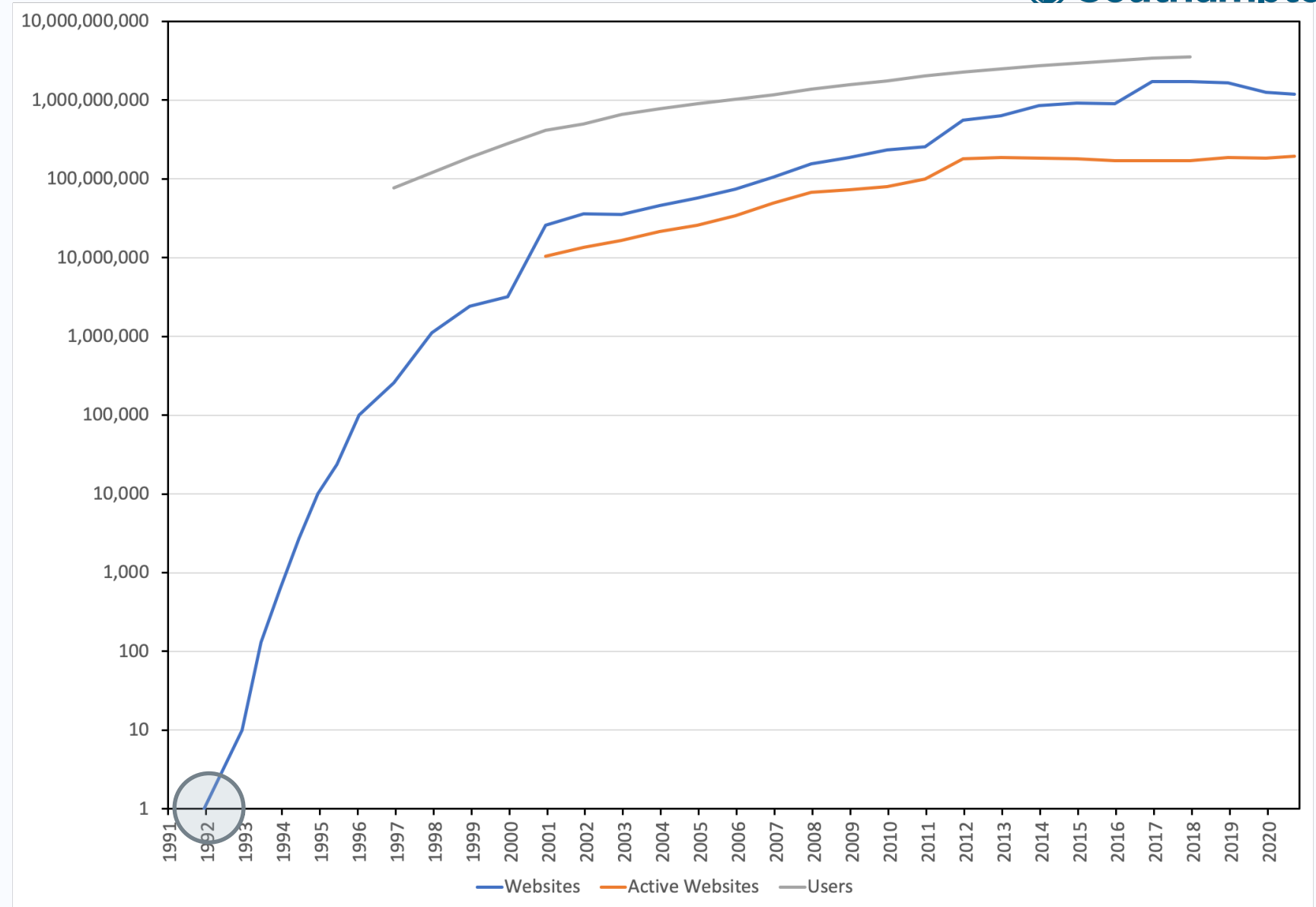
What is the World Wide Web

Web Growth

- Netcraft's survey of web servers
- Number of users - International Telecommunications Union
- Monthly survey of active websites

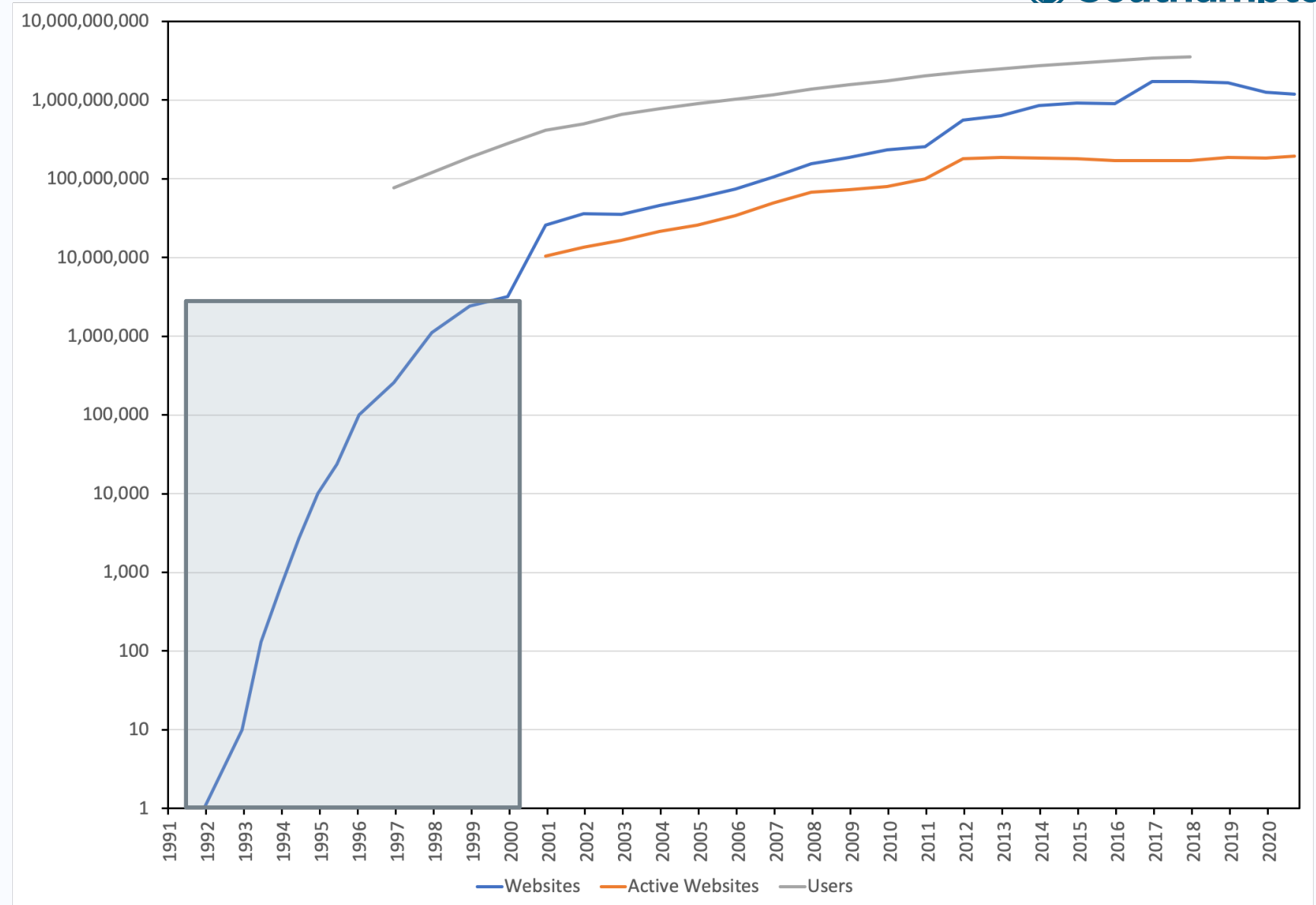


Web Growth

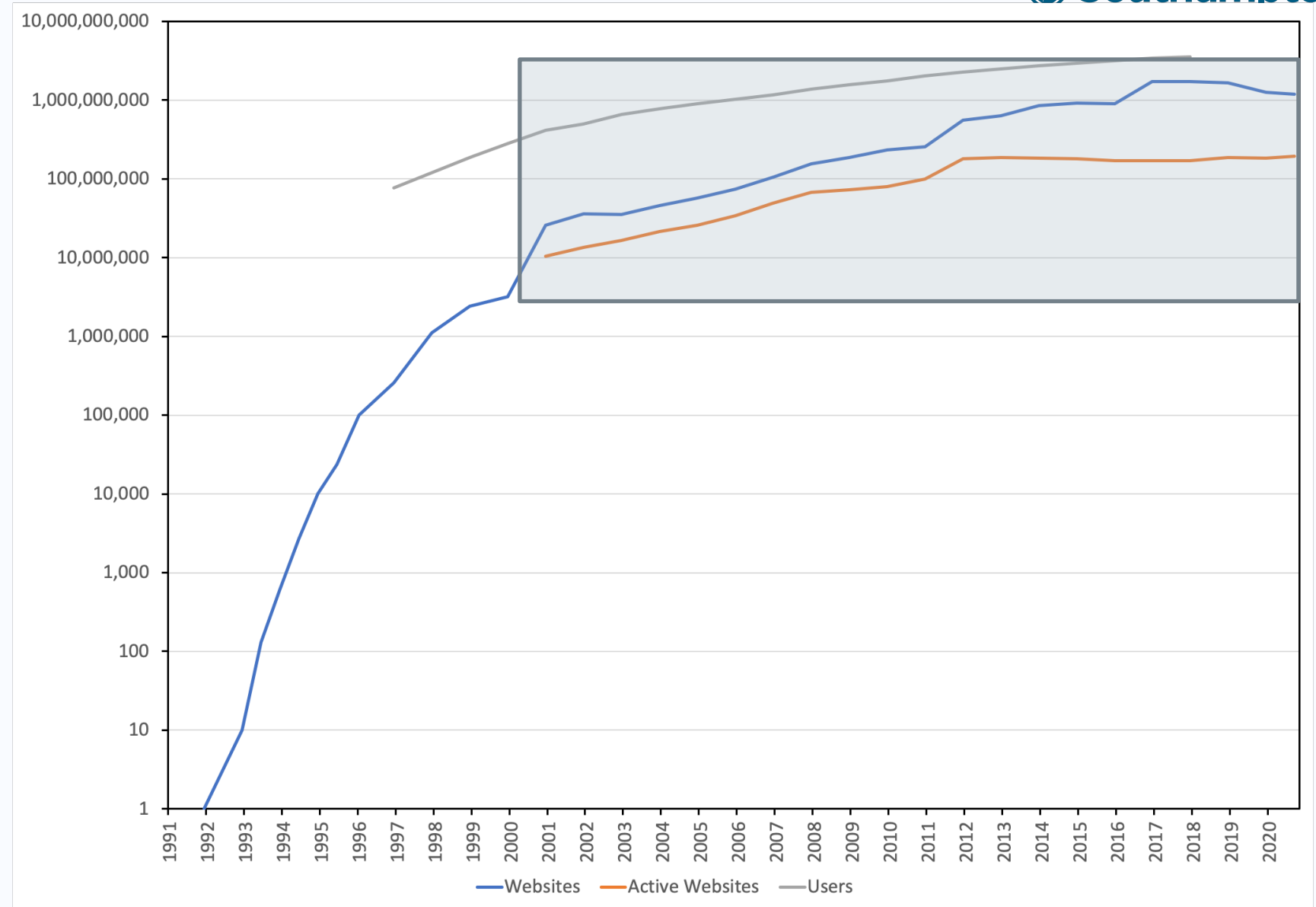


Web Growth

- Increasing by a factor of 10

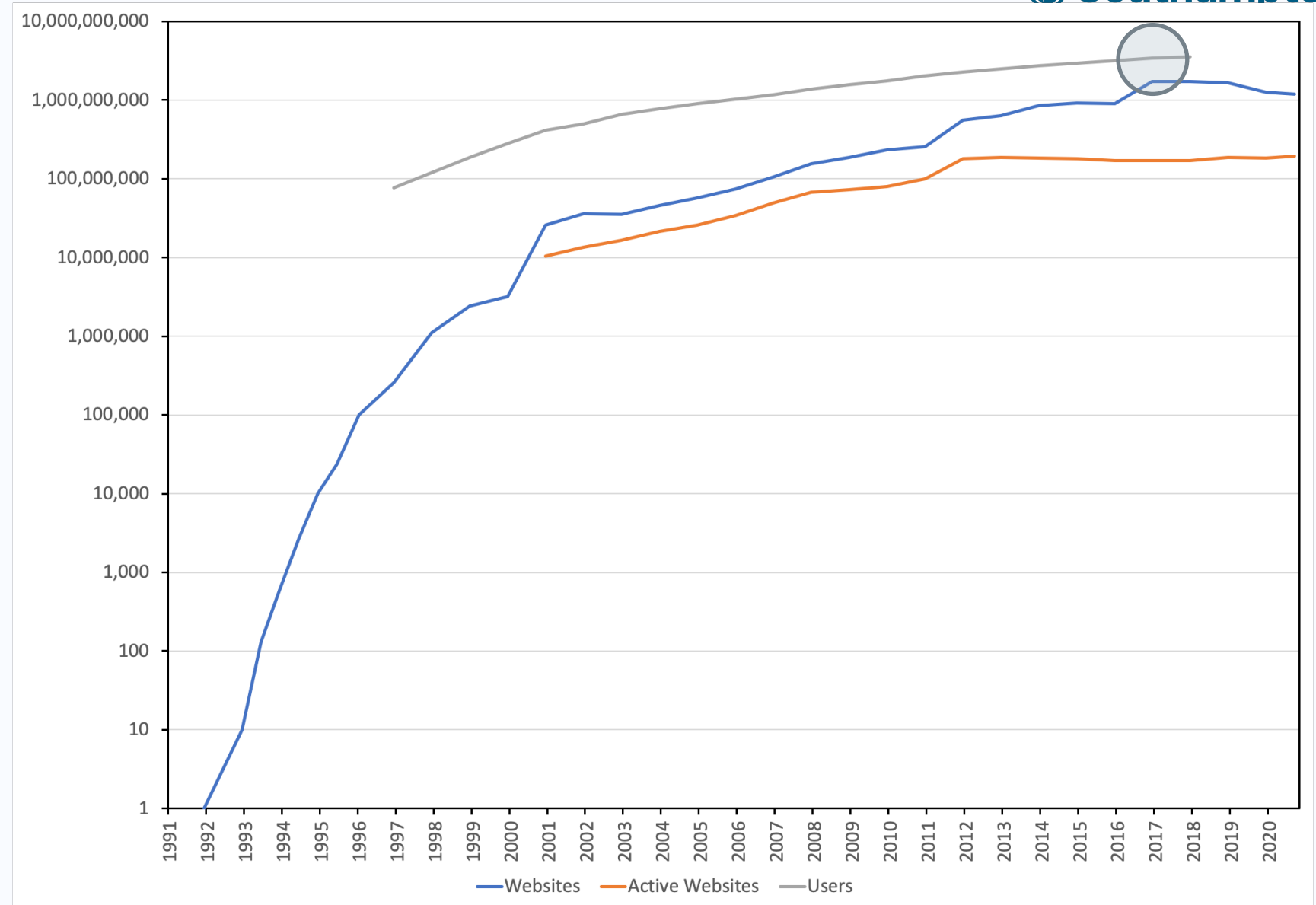


Web Growth



Web Growth

- 4.5 Billion Users 2017
- 7.5 Billion People 2017
- 8 Billion People 2022
- 100 different pages a day



Measuring the Web

How many webpages are there?

- Harder to count than websites (deep Web – not linked)
- Estimate size of the indexed Web using search engines
- Around 800 million in 1999 (compare with ~3 million websites)
- At least 10 billion in 2005 (compare with ~34 million active websites)
- At least 1 trillion in 2016 (compare with ~170 million websites)

What is the diameter of the Web?

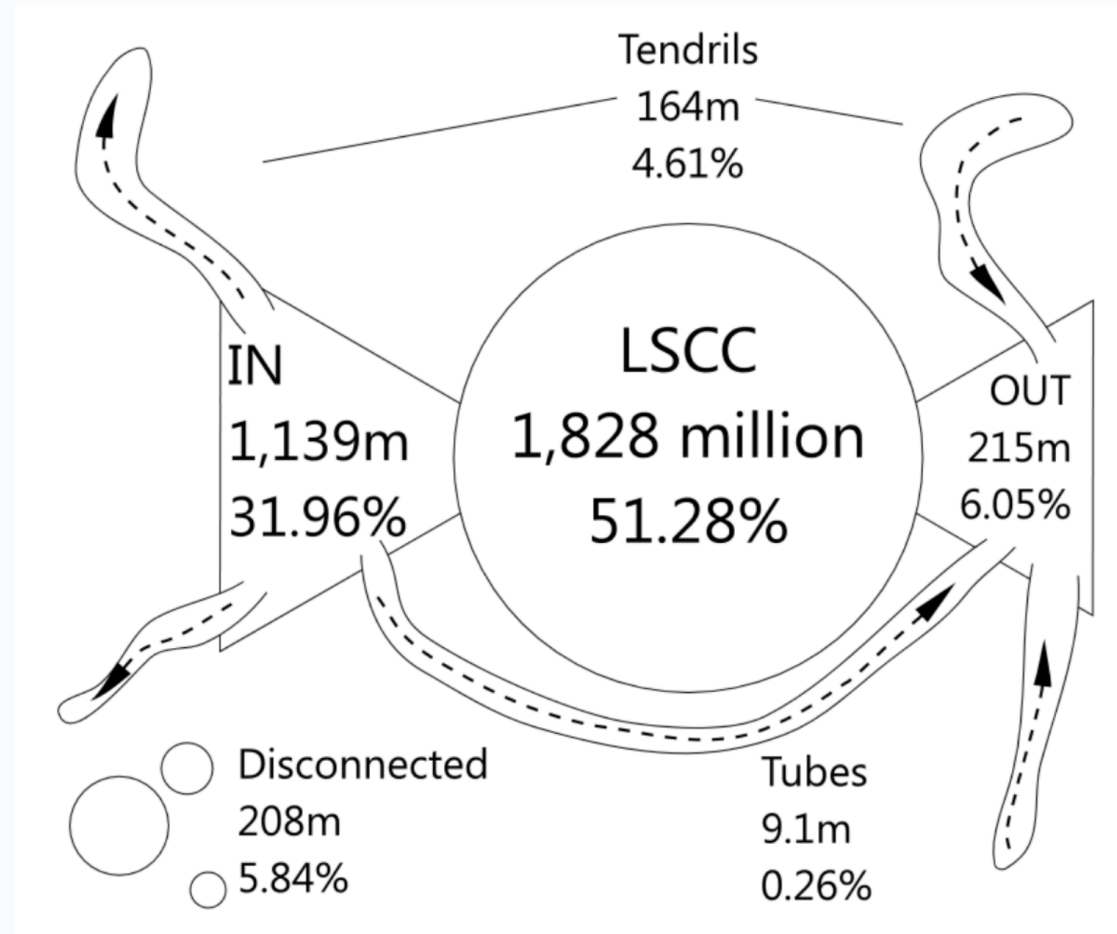
- How many links do you need to follow to travel between an arbitrary pair of webpages?
- Longest shortest finite path
- Even harder to measure...
- Estimated at 19 in 1999

Lawrence, S. and Giles, C.L. (1999) Accessibility of information on the Web. *Nature*, 400, pp. 107-109.

Albert, R. and Barabási, A-L. (1999) Diameter of the World-Wide Web. *Nature*, 401, pp.130-133.

The Shape of the Web

- Meusel et al. 2015
- LSCC - Large strongly connected component



Broder, A. et al (2000) *Graph Structure in the Web*. Computer Networks 33, pp. 309-320.

Meusel, R. et al (2015) *Graph Structure in the Web - Revisited*. In Proceedings of the 14th International World Wide Web Conference. pp. 427-432.

Next Lecture: Hypertext