

CC - Creative Commons Licensing

- This set of slides is part of the lecture „Semantic Web Technologies“ held at Karlsruhe Institute of Technology
 - The content of the lecture was prepared by PD Dr. Andreas Harth based on his book „Introduction to Linked Data“
 - The slides were prepared by Benedikt Köhler and Andreas Harth
-
- **This content is licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0):**
<http://creativecommons.org/licenses/by/4.0/>



Agenda

1. The Vision

1. Paul Otlet
2. Vannevar Bush
3. Doug Engelbart
4. Ted Nelson

2. The Internet

3. The World Wide Web

4. Towards a Web of Data



Paul Otlet

The arbitrary division into lines and pages of the book in its present format, does not correspond at all, with the presentation of ideas.

Hierarchical Modelling

- (4) *Europe (continent).*
- (49) *Petits pays de l'Europe (groupement de pays).*
- (493) *Belgique (pays).*
- (493.2) *Brabant (province).*
- (493.21) *Bruxelles (arrondissement).*
- (493.211) *Bruxelles (canton).*
- (493.211.1) *Bruxelles (ville).*

es choses

rs, la Réalité, le Cosmos

ntelligences

et les choses fragmentairement

La science

*une en ses cadres les pensées
ntelligences particulières*

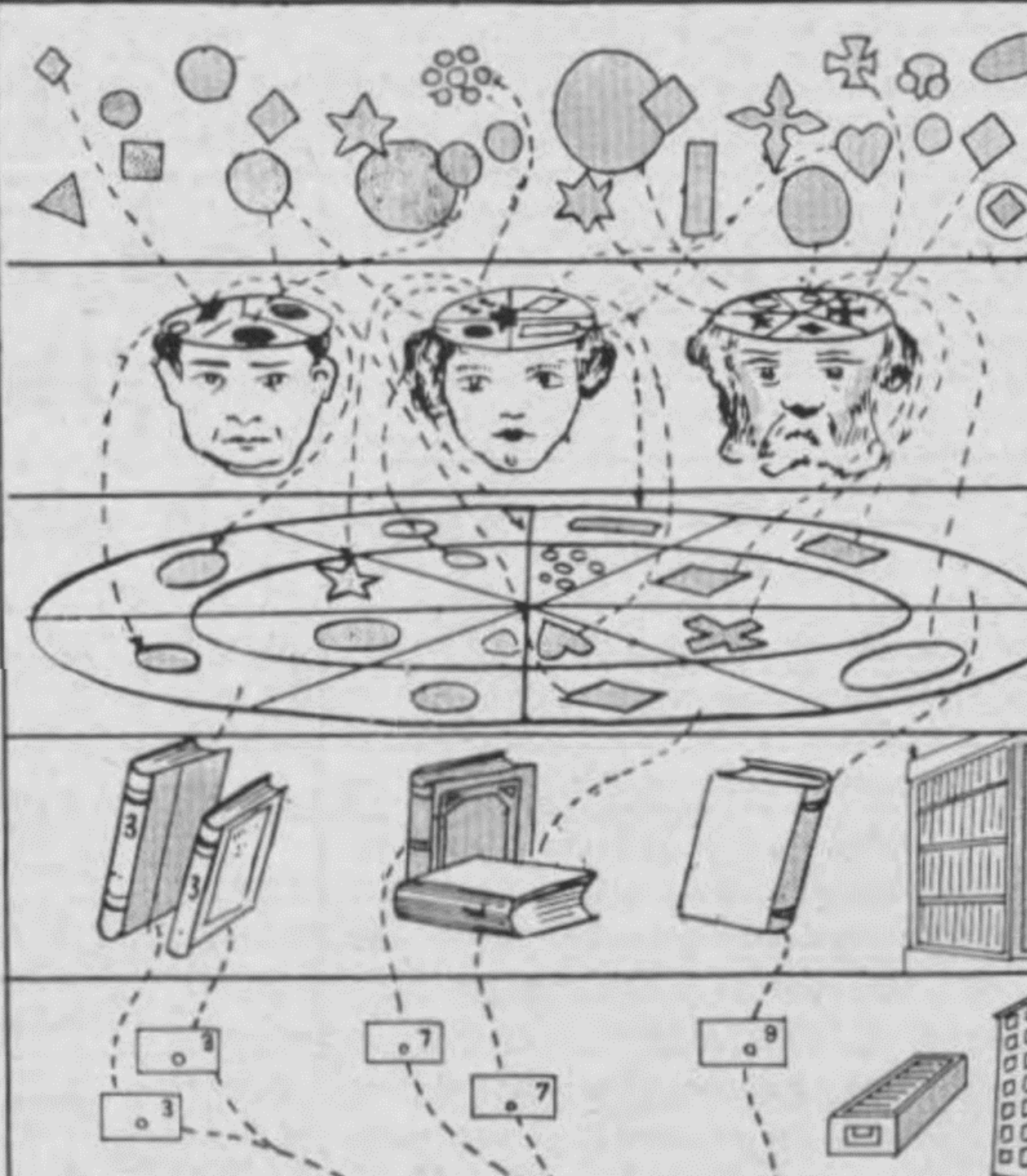
Les Livres

*et photographient la science
divisé des connaissances
de livres forment la Bibliothèque*

a Bibliographie

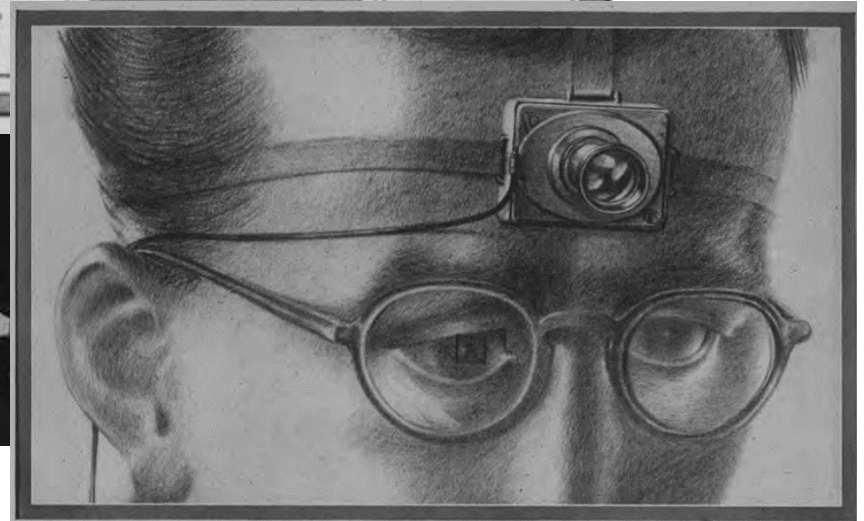
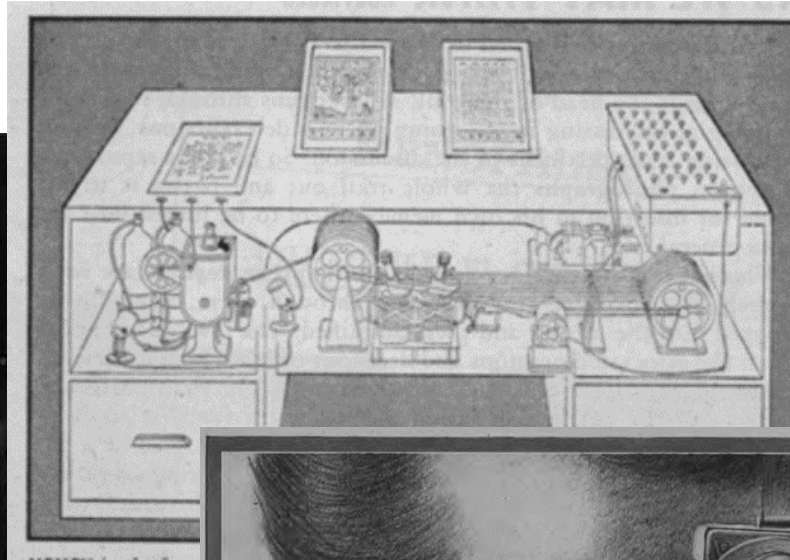
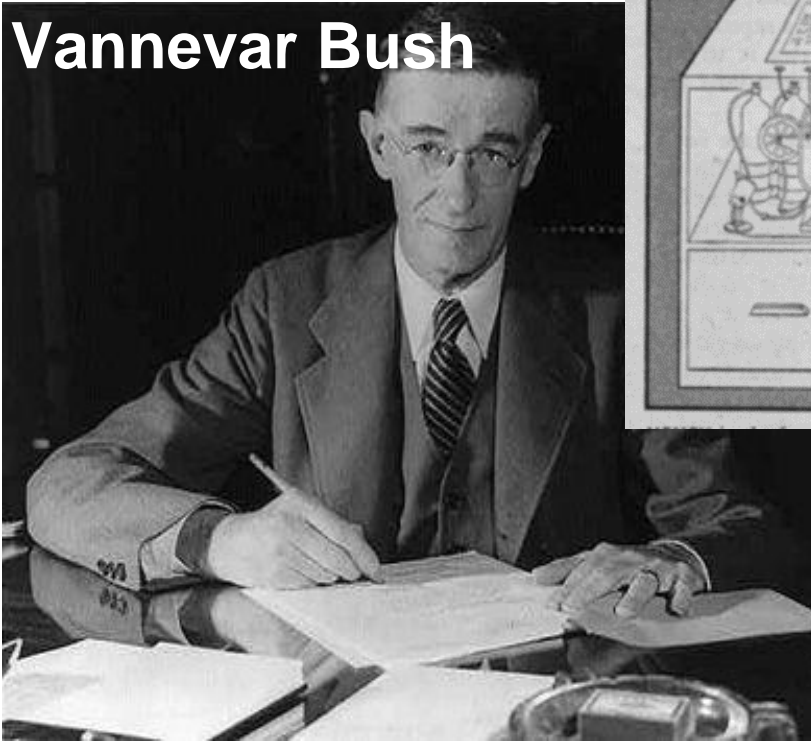
et catalogue les livres

notices Bibliographiques forme



The Memex (1945)

Vannevar Bush



The process of tying two items together is the important thing. . . . Thereafter, at any time, when one of these items is in view, the other can be instantly recalled merely by tapping a button . . .

Doug Engelbart

AUGMENTING HUMAN INTELLECT: A CONCEPTUAL FRAMEWORK

October 1962

Prepared for:

DIRECTOR OF INFORMATION SCIENCES
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
WASHINGTON 25, D.C.

CONTRACT AF 49(638)-1024

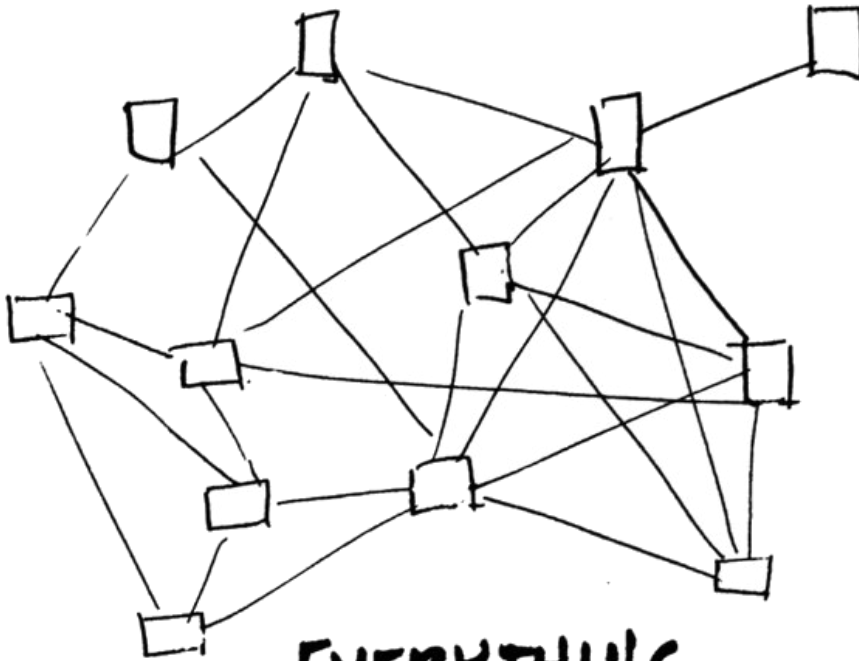
By: D. C. Engelbart

This is an initial summary report of a project taking a new and systematic approach to improving the intellectual effectiveness of the individual human being. A detailed conceptual framework explores the nature of the system composed of the individual and the tools, concepts, and methods that match his basic capabilities to his problems. One of the tools that shows the greatest immediate promise is the computer, when it can be harnessed for direct on-line assistance, integrated with new concepts and methods.

Ted Nelson

In an important sense there are no “subjects” at all; there is only all knowledge, since the cross-connections among the myriad topics of this world simply cannot be divided up neatly.

Hypertext at last offers the possibility of representing and exploring it all without carving it up destructively.



**EVERYTHING
IS DEEPLY INTERTWINGLED.**



Agenda

1. The Vision

2. The Internet

1. Vint Cerf and Bob Kahn
2. Internet Architecture
3. Internet Standardisation
4. The Network Effect

3. The World Wide Web

4. Towards a Web of Data

Vint Cerf and Bob Kahn



Internet Architecture and Standards

- Even though development and progress on the internet is achieved in a decentralised way, there is a need for governance and standardisation

Development Through RFCs

- RFCs are an example of how to manage standardisation in fast developing surroundings
- Proposed specifications have certain status categories¹ (e.g., Informational, Best Current Practice, Standards Track)
- Standards Track is divided into Draft Standard, Proposed Standard, and Internet Standards
- Internet Standards are specifications for which significant implementation and successful operational experience has been obtained²

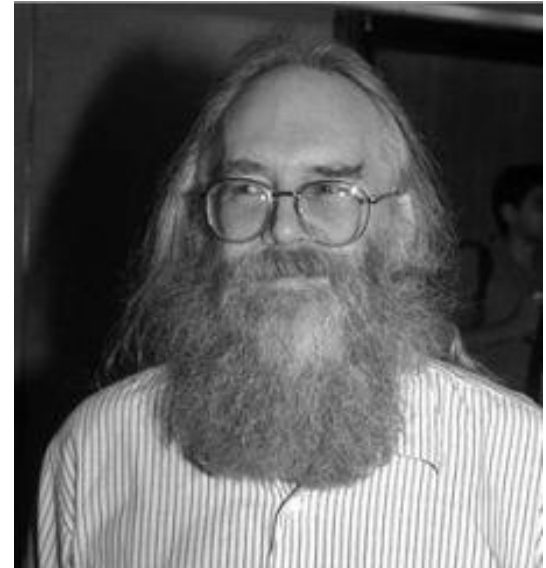
¹ <http://tools.ietf.org/html/bcp9#section-4>

² <http://tools.ietf.org/html/bcp9#section-4.1.3>

Postel's Robustness Principle

- Jon Postel (1943-1998): former RFC editor² and Internet Hall of Fame Pioneer
- „Robustness Principle“ (RFC 793)

*“...be conservative in what you do,
be liberal in what you accept from others.”³*



1

- Meaning: Internet systems should be strict in what they emit, but liberal in what they accept

¹ <http://www.internethalloffame.org/inductees/jon-postel>

² Posthumously, he was honored with his own RFC where he is described as leader, icon, giant and „Internet Assigned Numbers Authority“ (<https://tools.ietf.org/html/rfc2468>)

³ <https://tools.ietf.org/html/rfc793#section-2.10>

Internet Engineering Task Force

- The Internet Engineering Task Force (IETF) is hosted at an organisation called the Internet Society.
- The IETF provides a forum for interested parties to work on specifications for internet technology specifications
- These specifications are called Request for Comments (RFCs) and describe a toolbox of protocols, methods and concepts that are essential for operating systems on the internet
- The IETF also published RFCs that serve as foundation of the World Wide Web: Uniform Resource Identifiers (URIs) and Hypertext Transfer Protocol (HTTP)



¹ <http://ietf.org/>

Internet Corporation for Assigned Names and Numbers

- The Internet Corporation for Assigned Names and Numbers (ICANN) is the main organisation that governs the internet



- ICANN is:²

An internationally organised, non-profit corporation with responsibility for

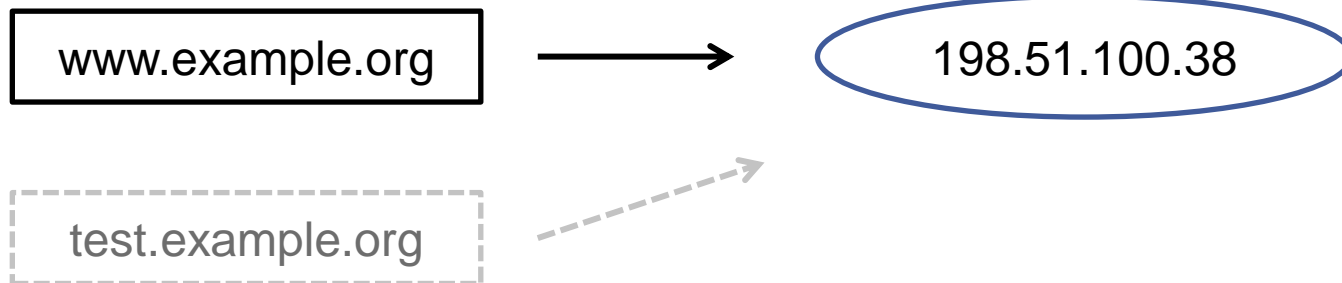
- Internet Protocol address space allocation,
- protocol identifier assignment,
- generic and country code top-level Domain Name System management,
- and root server system management functions.

¹ <https://archive.icann.org/images/icannlogo.jpg>

² <https://www.icann.org/en/system/files/files/participating-08nov13-en.pdf>

Internet Assigned Numbers Authority

- The Internet Assigned Numbers Authority (IANA) is a department of ICANN that
 - coordinates the allocation of IP addresses and
 - administers the Domain Name System (DNS)
- The DNS is hierarchically organised and provides means to map from hostnames to IP addresses

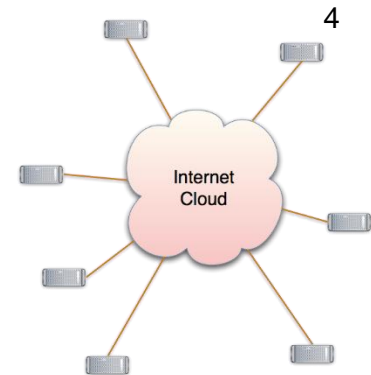


Multiple hostnames can map to the same IP!

¹ <http://iana.org/>

Examples of IETF Internet Standards

- Internet Protocol (IP) – RFC 791¹ (Sept 1981)
Responsible for sending packets to devices in a network
- Transmission Control Protocol (TCP) – RFC 793² (Sept 1981)
Responsible for breaking data into packets to fit IP
- Domain Names – Concepts and Facilities – RFC 1034³ (Nov 1987)
Mapping hostnames to IP addresses



¹ <http://tools.ietf.org/html/rfc791>

² <http://tools.ietf.org/html/rfc793>

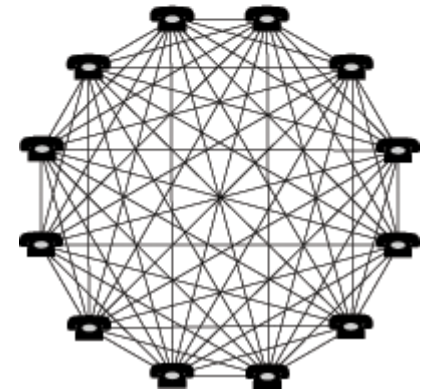
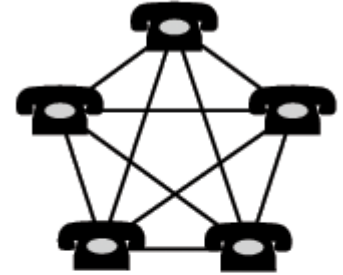
³ <http://tools.ietf.org/html/rfc1034>

⁴ <http://www.w3.org/DesignIssues/Abstractions.html>

The Network Effect aka Metcalfe's Law

“According to Metcalfe, the rationale behind the sale of networking cards was that (1) the cost of the network was directly proportional to the number of cards installed, but (2) the value of the network was proportional to the square of the number of users. This was expressed algebraically as having a cost of N , and a value of N^2 .”

http://en.wikipedia.org/Network_effect



Walled Gardens and the Platform Economy

- Centralised information systems that facilitate social and economic interactions
- Social interactions: Facebook
 - Business model: run advertisements based on user data
- Economic interactions: Alibaba
 - „Transaction platforms“ act as centralised matchmakers between demand and supply
 - Business model: take a fee on transactions; upsell services to participants; mine and monetise the data trail of participants
- The Internet and the World Wide Web also facilitate social and economic interactions, but without a company running and monetising the centralised platform



The Economist, Illustration by David Simonds

Agenda

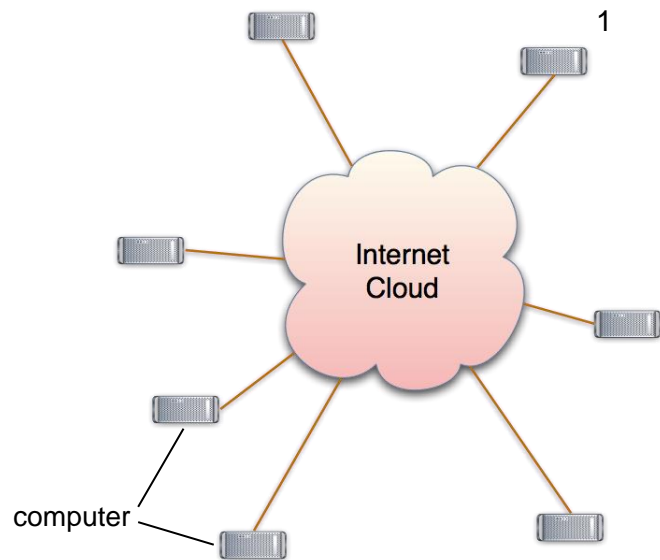
1. The Vision
2. The Internet
- 3. The World Wide Web**
 1. Tim Berners-Lee
 2. Web Architecture
 3. Resources and Uniform Resource Identifiers (URIs)
 4. The Hypertext Transfer Protocol (HTTP)
 5. The Hypertext Markup Language (HTML)
 6. User Agent Loop
 7. User Agents and Servers
 8. The World Wide Web Consortium
4. Towards a Web of Data

Tim Berners-Lee

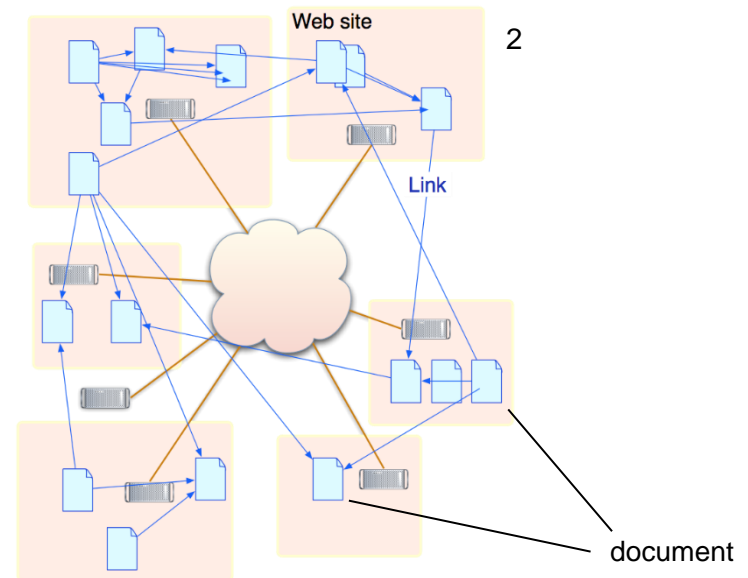
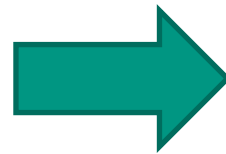


Web Architecture and Standards

- The so far presented Internet Architecture is the foundation of the World Wide Web
- Let's have a quick look on the organisation and technologies of the web!



Internet Model



WWW Model

^{1,2} <http://www.w3.org/DesignIssues/Abstractions.html> (great wrap-up of the developing understanding of the internet & web)

World Wide Web Consortium W3C



1

“The World Wide Web Consortium (W3C) is an international community where Member organizations, a full-time staff, and the public work together to develop Web standards.”²

- W3C aims for the worldwide availability of web access and envisions web progress in terms of interaction, data, services, and security through standardization³
- Tim Berners-Lee is the director of W3C

¹ <http://www.w3.org/>

² <http://www.w3.org/Consortium/>

³ <http://www.w3.org/Consortium/mission>

W3C Drafts and Recommendations

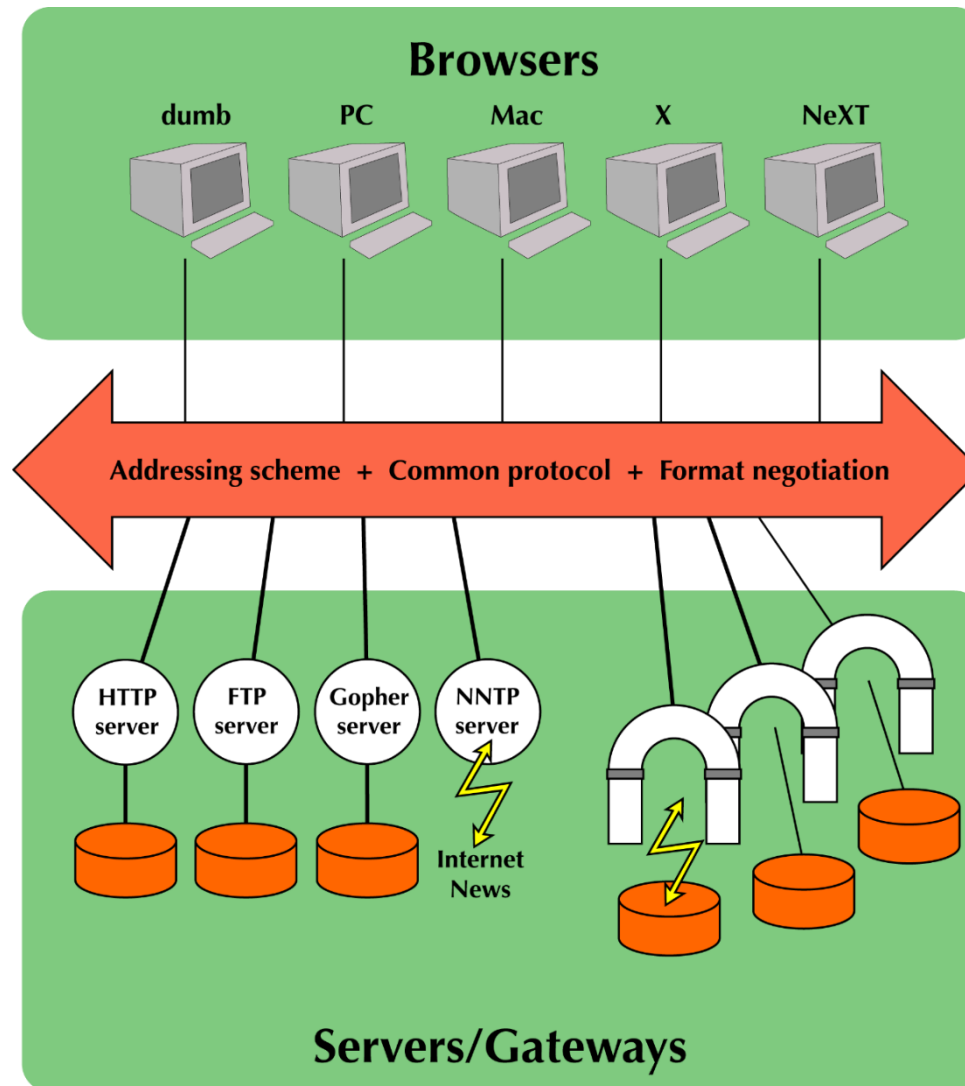
- Comparable to IETF's standardisation through RFCs, W3C publishes reports that pass different maturity levels until they are officially recommended¹:
 1. Working Draft (WD)
 2. Candidate Recommendation (CR)
 3. Proposed Recommendation (PR)
 4. W3C Recommendation (REC)

- In this lecture, we are concerned with languages specified in W3C recommendations²

¹ <http://www.w3.org/2005/10/Process-20051014/tr.html#maturity-levels>

² <http://www.w3.org/TR/>

Web Architecture Overview



<https://www.w3.org/People/Berners-Lee/1996/ppf.html>

Architecture of the World Wide Web

- The Web Architecture Group at W3C is concerned with the principles of web architecture and published the document „ Architecture of the World Wide Web, Volume One”¹ in 2004
- The document covers, amongst other topics, the correct use of URIs and HTTP in context of the web
- The theoretical foundation for web architecture principles is described as Representational State Transfer (REST)²

¹ <http://www.w3.org/TR/webarch/>

² Chapter 5 of "Architectural Styles and the Design of Network-based Software Architectures", Doctoral Thesis of R. T. Fielding, 2000. http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm

Resources

- Resources are anything we want to talk about
- Uniform Resource Identifiers (URIs) denote („are names for“) resources

- The following are resources:

- electronic documents
- images
- collections of other resources
- chemical elements
- numeric values
- rooms

} “files”: Information Resources

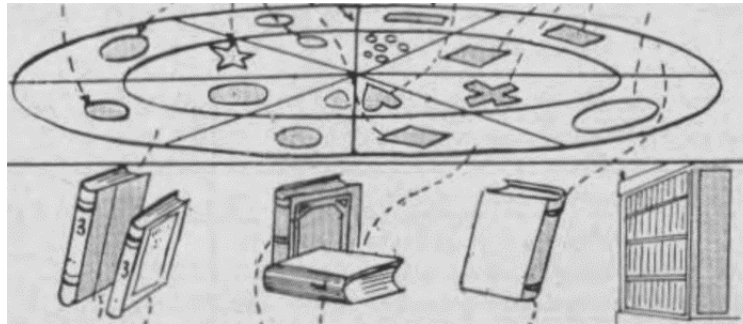
} “directories”: Collection Resources

} But also abstract concepts!

} “Non-information Resources”

Important Distinction

- We need to distinguish between other resources (Otlet's „frames of thought) and information resources (Otlet's books)



- We also need to distinguish between the name of a thing (the URI) and the thing itself (the resource)

Uniform Resource Identifiers

- A Uniform Resource Identifier (URI) is a character sequence that identifies a resource.
- URIs are defined in RFC 3986 (Jan 2005; Internet Standard)



Room sign at WU Vienna with HTTP URI. A lookup on the room's URI redirects to a web page which allows users to view the room schedule and book the room.

¹ <http://tools.ietf.org/html/rfc3986>

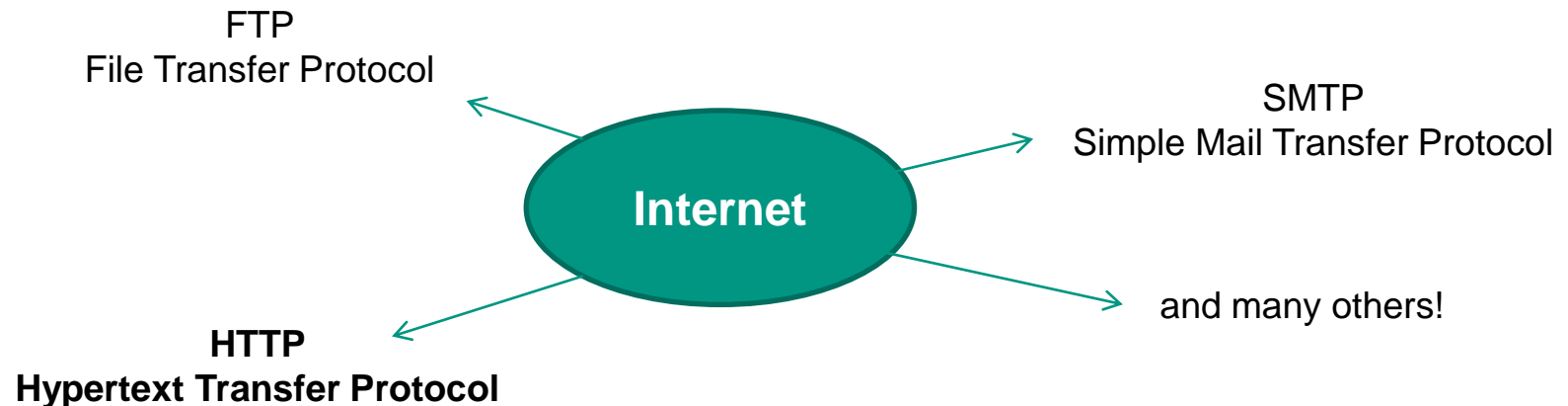
Examples of Valid URIs¹

- `ftp://ftp.is.co.za/rfc/rfc1808.txt`
- `http://www.ietf.org/rfc/rfc2396.txt`
- `ldap://[2001:db8::7]/c=GB?objectClass?one`
- `mailto:John.Doe@example.com`
- `news:comp.infosystems.www.servers.unix`
- `tel:+1-816-555-1212`
- `telnet://192.0.2.16:80/`
- `urn:oasis:names:specification:docbook:dtd:xml:4.1.2`

¹ <http://www.rfc-editor.org/rfc/rfc3986.txt>

Protocols

- The internet is the networking infrastructure for global data exchange
- Information is accessible on the internet via different protocols
- HTTP (Hypertext Transfer Protocol, foundation for the World Wide Web) is one of the protocols



Hypertext Transfer Protocol - Definition

- RFC 7230¹ (Standard) states:

The Hypertext Transfer Protocol (HTTP) is a stateless application-level protocol for distributed, collaborative, hypertext information systems.

- A user agent poses a request to a server, and a server issues a response to the user agent
- User agents can be web browsers or command line programs running on PCs or mobile phones, but also software components that run on household appliances, cars, internet radio devices...
- Servers can be traditional large-scale web servers of commercial companies, but also software components that offer access to sensor data or actuator functionality (on household appliances, cars,...)

¹ <http://tools.ietf.org/html/rfc7230>

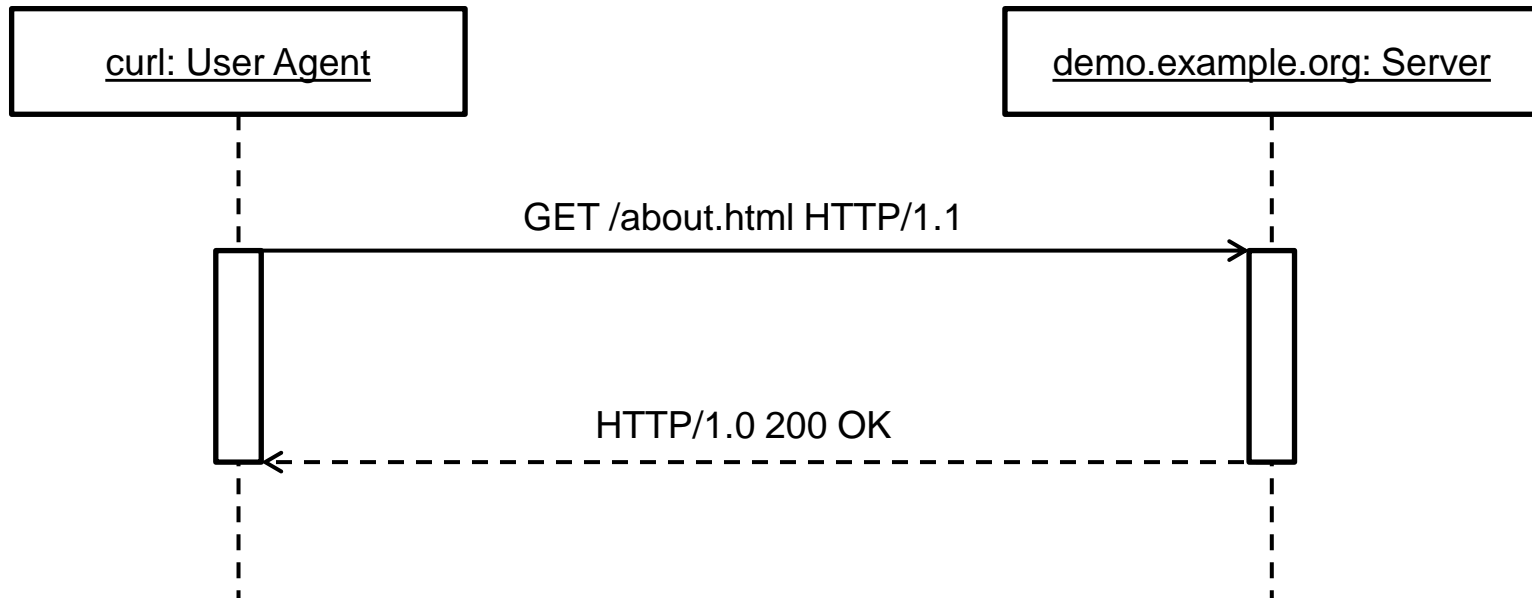
Successful Request/Response Interaction

- A simple successful request/response interaction between user agent and server works as follows:
 1. The user agent establishes a TCP connection with the server.
 2. The user agent transmits a HTTP request. Trailing newline to indicate end of message.
 3. The server replies with a HTTP response. Trailing newline to indicate end of message.
 4. The connection is terminated.

HTTP Requests and Responses

- A HTTP request consists of
 - a HTTP request method (e.g., GET, PUT, POST, DELETE), the request URI, and HTTP protocol version information
 - optionally a list of HTTP headers consisting of name/value pairs
 - optionally a message body
- A HTTP response consists of
 - HTTP protocol version information and a HTTP status code
 - optionally a list of HTTP headers consisting of name/value pairs
 - optionally a message body

UML Sequence Diagram



HTTP Status Codes (RFC 7231)

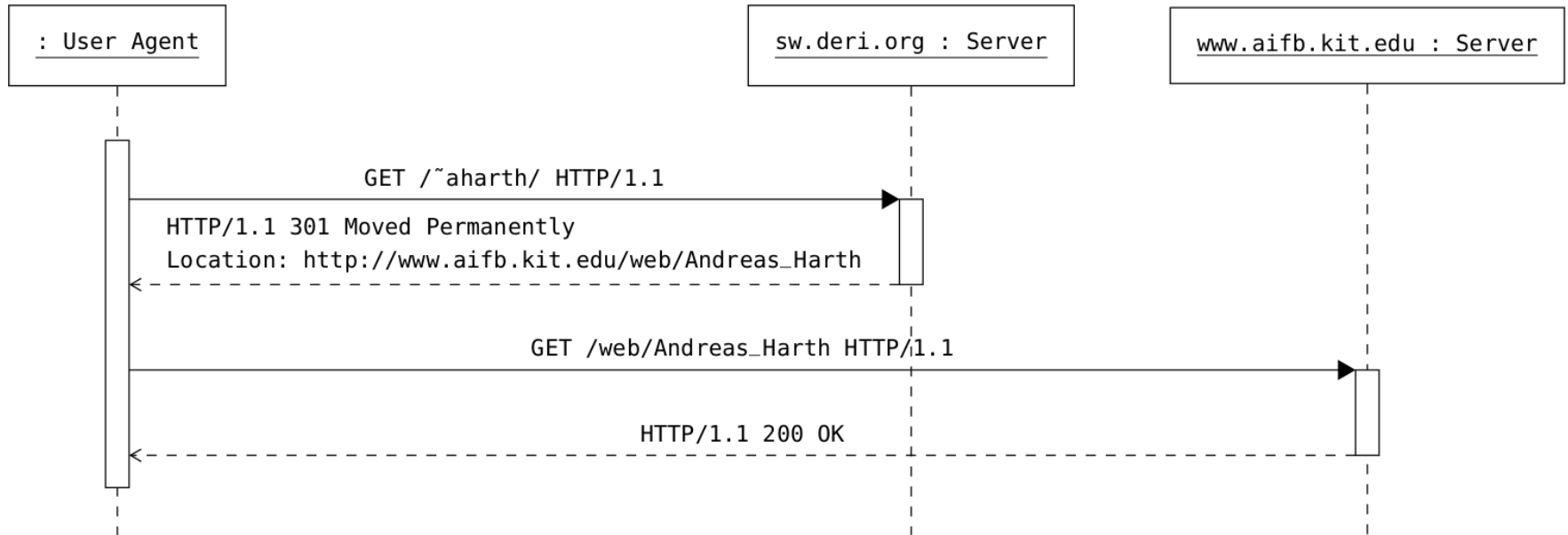
- By status codes, the server interacts with the user agent, reporting success, failure, etc.
- The status code is returned in the start line of the response message
- You know the 404 Not Found status code

```
HTTP/1.0 200 OK
Server: Apache
Content-Type: text/html
```

```
<html><body><h1>Demo
</h1></body></html>
```

Status code classification		
1xx	Informational	provisional response
2xx	Successful	request successfully received, understood, and accepted
3xx	Redirection	further action needs to be taken by user agent to fulfil the request
4xx	Client Error	client erred
5xx	Server Error	server encountered an unexpected condition

HTTP Redirects



- Servers can tell the user agents that the location of resources has changed via the 3xx status codes.
- The Location header specifies the new URI of the requested resource.

Content Negotiation¹

- Typically, the user-agent has a preference about the content-type of the requested data (format and encodings).
- The preference is indicated in the Accept header of the HTTP request, e.g. the header `Accept: image/jpg, image/png` indicates that the user agent can handle JPEG and PNG images
- The server replies with the specified format of the requested file; the format is indicated via the Content-Type header.
- Other headers for content negotiation include the Accept-Language header, in which the user agent advertises the preferred languages, e.g., `Accept-Language: de`.

¹ <http://tools.ietf.org/html/rfc7231#section-3.4>

Hypertext Markup Language HTML

- HTML can be used to encode Hypertext documents
- In 1995, HTML 2.0 was specified as an IETF RFC 1866¹
- The next version, HTML 3.2², reached W3C recommendation status in 1997
- HTML 5 was initiated in 2004 by individuals of Apple, the Mozilla Foundation and Opera Software, who were concerned with the direction the W3C was taking with XHTML (HTML based on stricter XML syntax rules)
- Ultimately, HTML 5 was finalised within the W3C process³

¹ <http://tools.ietf.org/html/rfc1866>

² <http://www.w3.org/TR/REC-html32/>

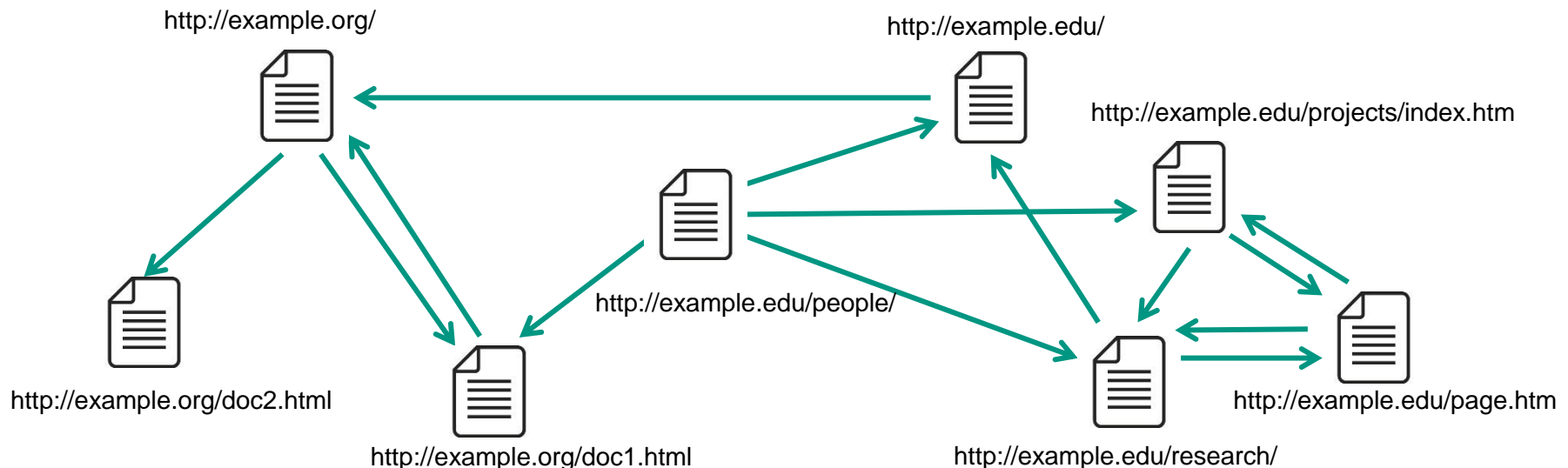
³ <http://www.w3.org/TR/html5/>

Hypertext Markup Language HTML

- HTML can be used to represent document structure, such as headings and paragraphs
- Further, documents can reference other documents (e.g., via `a href`), include media (via `img src`) or provide a form (via `form action`)
- With HTML, we could fill several lessons!

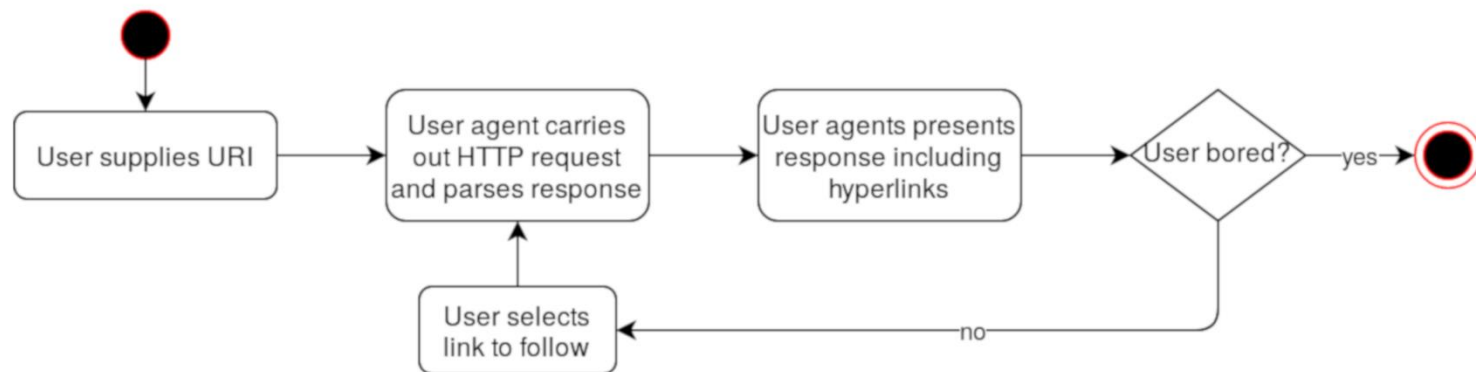
Hypertext Markup Language HTML

- For the purpose of the lecture, the most important feature of HTML is the possibility to link to other documents!
- A user agent (web browser) uses links to enable users to navigate to other pages and to display additional information



User Agent Loop

- The User Agent Loop is a model to describe how a user interacts with a user agent to 'surf' on the Web



User Agent Loop

1. The user agent starts its interaction based on a user supplied URI

2. The user agent performs HTTP requests on URIs and parses the response

3. Based on the response the user agent has one or multiple choices as to which interaction to perform next

4. The user agent decides which link to follow and initiates a new request, go to 2.

Compact URIs (CURIEs)

- We will work a lot with URIs, but full URIs can be unwieldy
- Thus, there is a syntax for abbreviated URIs¹ called Compact URIs, or CURIEs for short²
- CURIEs consist of a prefix (“namespace”) and a local reference (“local part”)
- Assume we declare the prefix `abc` with a value of `http://example.org/doc.ttl#`
- With the prefix `abc` declared, the CURIE `abc:Berlin` expands to

`http://example.org/doc.ttl#Berlin`

¹ <http://www.w3.org/TR/curie/>

² CURIEs are an extension to QNames, which are used to abbreviate attribute URIs in XML documents

URIs in Relative Form

- In contrast to absolute HTTP URIs (those starting with `http://` and including a hostname), HTTP URIs can also occur in relative form
- They have to be interpreted *relatively* to an absolute URI
- A URI-reference is either a URI or a relative reference¹
- We can also use the notation known from file systems: “.” refers to the current directory, while “..” refers to the parent directory²

Relative reference	Base URI	Resolves to the URI
research/	<code>http://example.edu/</code>	<code>http://example.edu/research/</code>
<code>./academics/</code>	<code>http://example.edu/research/</code>	<code>http://example.edu/research/academics/</code>
<code>../academics/</code>	<code>http://example.edu/research/</code>	<code>http://example.edu/academics/</code>
<code>#people</code>	<code>http://example.edu/research/</code>	<code>http://example.edu/research/#people</code>
	<code>http://example.edu/doc</code>	<code>http://example.edu/doc</code>

¹ <http://tools.ietf.org/html/rfc3986#section-4.1>

² for detailed technical instructions and further examples: <http://tools.ietf.org/html/rfc3986#section-5.2>

Agenda

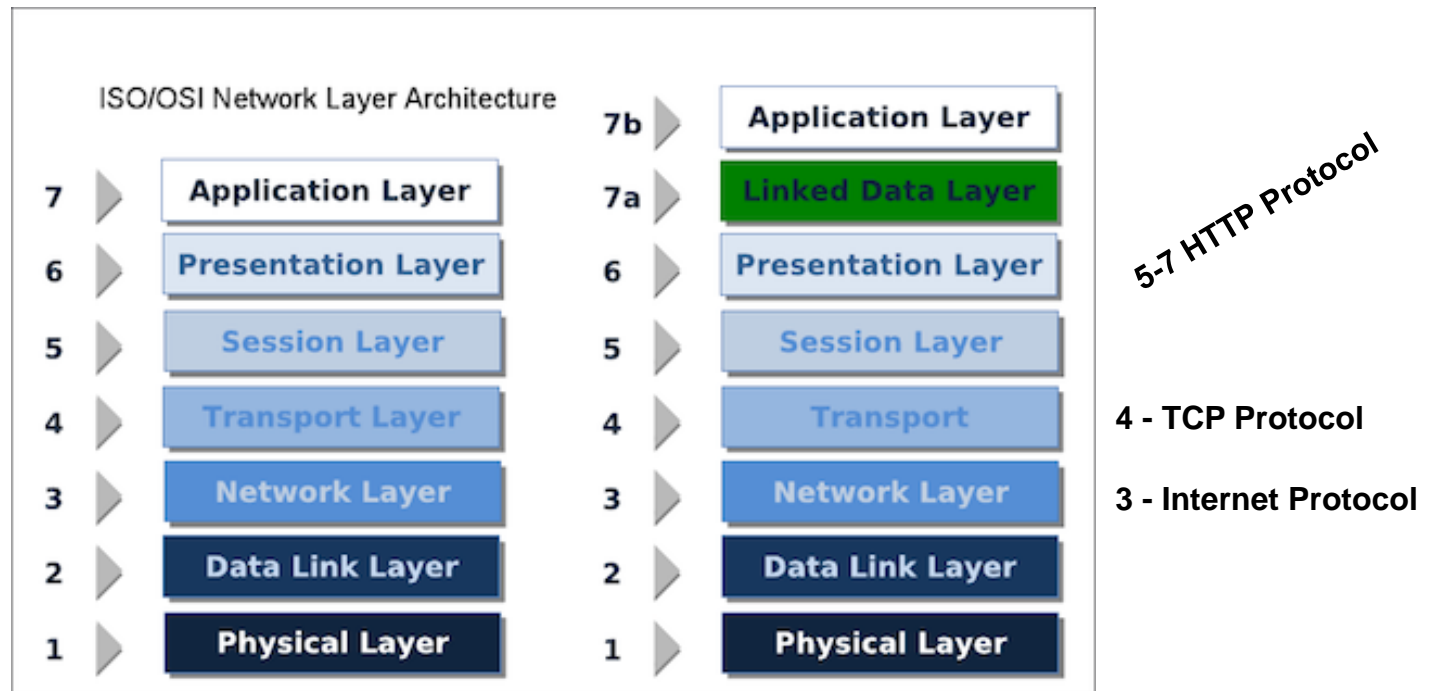
1. The Vision
2. The Internet
3. The World Wide Web
4. **Towards a Web of Data**

A Web for Machines

- We would like to use web architecture as basis for machines accessing information and carrying out tasks for their users
- In other words, we would like to have user agents that users can delegate tasks to (for example, access multiple sources, integrate the data and run a query over the combined data)
- We require a way to access (arbitrary) data and the ability to integrate arbitrary data
- Thus, the data has to be self-describing, for anyone to pick up and understand

Towards a Semantic Web architecture

- The Open Systems Interconnection (OSI) model splits communication functions into seven layers
- Moving from the Web of Documents to a Web of Data, the Linked Data Layer is introduced that focuses on ensuring data interoperability



Source: <https://technologyvoice.com/2010/09/06/the-internet-becomes-the-interdata-interview-with-stefan-decker/>

Extensible Markup Language (XML)

- XML is a data-exchange format
- XML documents must adhere to stricter syntactic rules than traditional HTML
- The stricter model facilitates for machine processing of XML documents
- Attributes in XML do use URIs for identification, however, XML does not provide the capability to reference URIs as attribute values¹
- XML 1.0 became W3C recommendation in 1998²

```
<corps>  
  <prenom>Paul</prenom>  
  <nom-de-famille>Otlet</nom-de-famille>  
  <lieu>Bruxelles</lieu>  
</corps>
```

¹ S. Decker, S. Melnik, F. van Harmelen, D. Fensel, M. Klein, J. Broekstra, M. Erdmann, and I. Horrocks. The semantic web: the roles of xml and rdf. Internet Computing, IEEE, 4(5): 63–73, Sep 2000. <http://doi.org/10.1109/4236.877487>

² <http://www.w3.org/TR/1998/REC-xml-19980210>

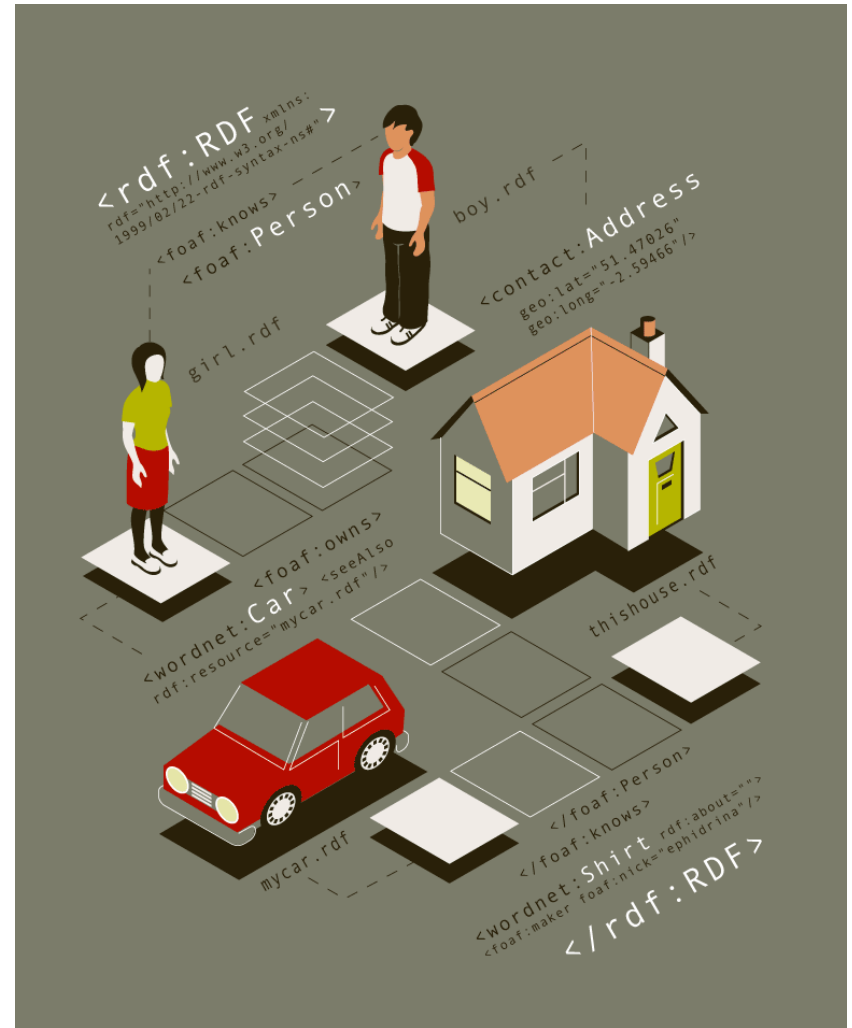
JavaScript Object Notation (JSON)

- JSON is a data-interchange format for tree-structured data
- JSON documents consist of Numbers, Strings, Booleans, Arrays and Name-Value Pairs
- Standardised in an RFC (via the IETF)
- In the basic version no use of URIs

```
{  
  "nombre de pila": "Paul",  
  "apellido": "Otlet",  
  "lugar": "Bruselas"  
}
```

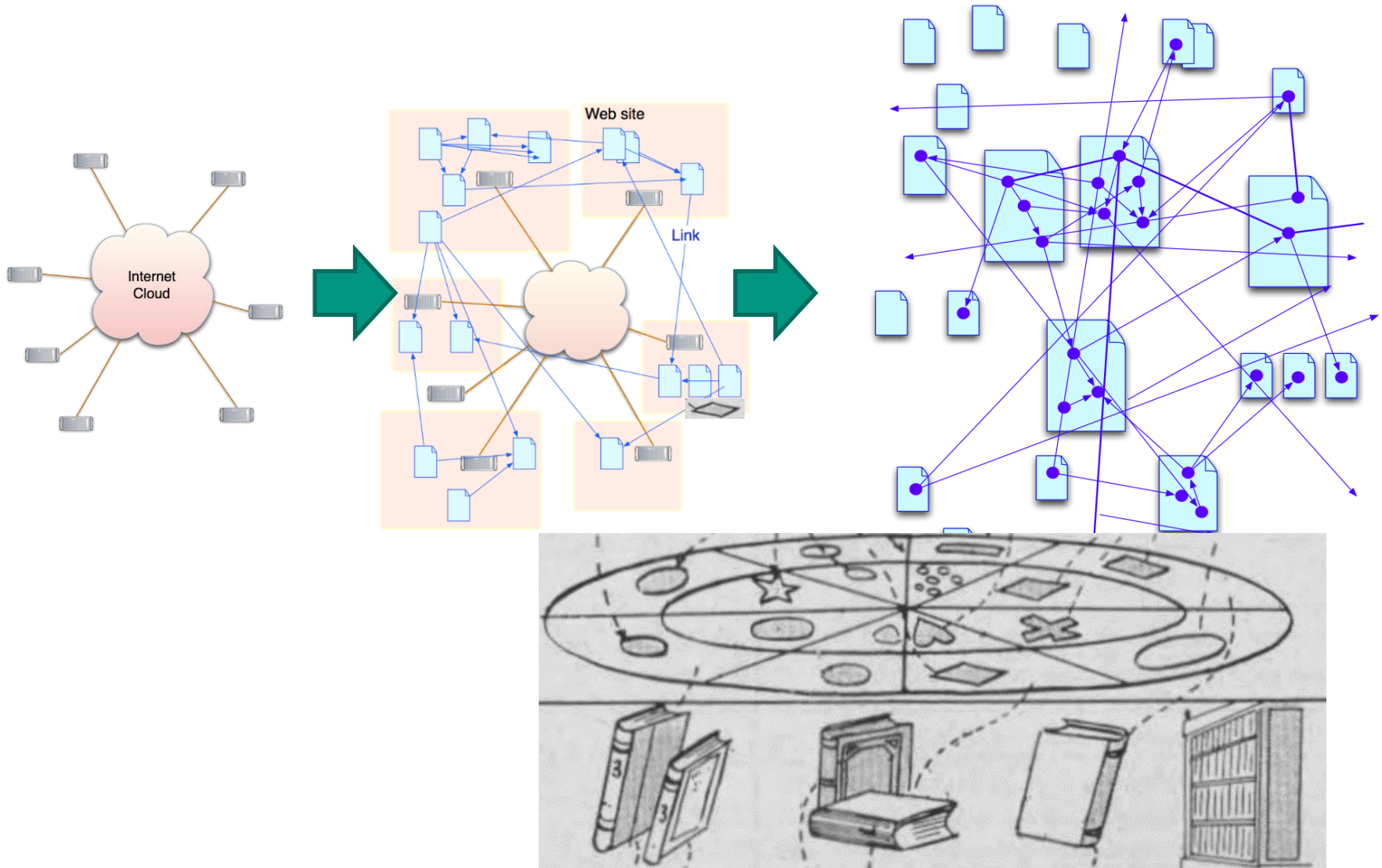
Think-Pair-Share: URIs for Cars and People

- What URIs would you give to yourself?
- Your car?
- Your house?



SWAD-Europe project website
<http://www.w3.org/2001/sw/Europe/>

Towards a Web of Data

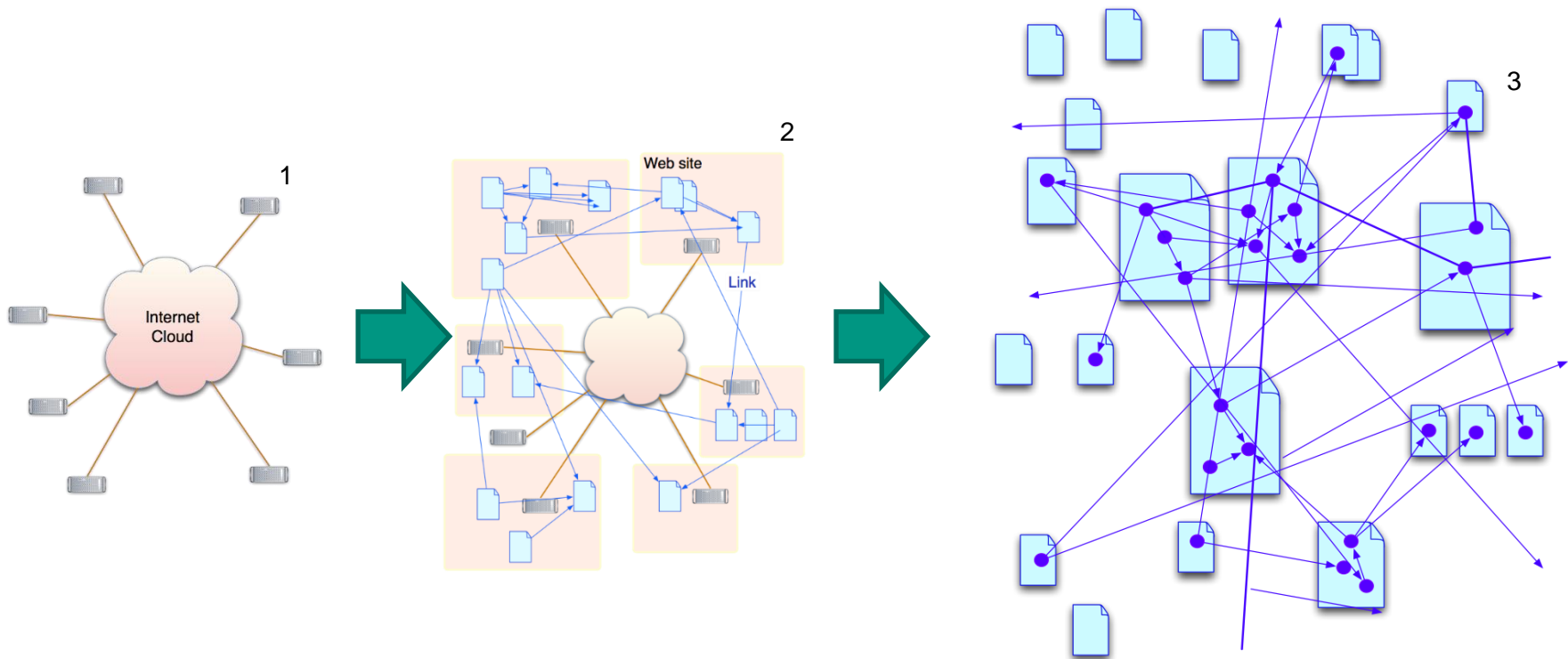


Learning Goals

- G 1.1 Distinguish between hierarchical structure and network structure in representing information.
- G 1.2 Outline basic internet and web technologies, describe the relation between them, and point out their differences.
- G 1.3 Assess the characteristics of decentralisation in information system architectures, and explain the network effect as expressed in Metcalfe's Law.
- G 1.4 Identify resources and name resources via URIs: in absolute form, in relative form, as CURIEs.
- G 1.5 Read and interpret HTTP request/response interactions between user agent and server, including HTTP redirects and content negotiation.

Outlook – Chapter 2

- In the next lecture, we will get to know the Semantic Web and the ideas behind Linked Data



1,2,3 <http://www.w3.org/DesignIssues/Abstractions.html>

Get the Idea?

(TOMORROW'S WEB WILL)



1 0 1 0 1 0 1 0 1
0 1 1 0 1 0 0 0 1
1 0 k n o w 0 1 0
0 w h a t 0 0 1 0
0 1 0 y o u 1 0 1
1 0 1 m e a n ...
1 0 1 1 0 0 1 0 0

PLUS:

Antibiotics'
Dim Future

Rorschach:
A Waste of Ink

The Oldest Stars