Source: http://lod-cloud.net

Semantic Web

Welcome!

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Prof. Dr. Stefan Decker

Sascha Welten, M.Sc.

Md. Rezaul Karim, M.Eng.

Lectures

Exercises

Contact via:
welten@dbis.rwth-aachen.de or **rezaul.karim@rwth-aachen.de**

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Information about the Exercises

- Exercises are mandatory, your need at least 50% points to qualify for the final exam.

- There will be 6 exercises: to be submitted online and to be auto-graded.

- Exercises (except for exercise 1) will be based on Jupyter notebooks.

- Notebooks can be opened via RWTH Jupyter Hub: https://jupyter.rwth-aachen.de

- Information about the exercises can be found in the Moodle room.

- Refer to "*Introduction to Jupyter Notebooks"*, which explains:

    - interaction between notebooks & Jupyter Hub.
    - a test upload.

1. Introduction

Introduction to Jupyter Notebooks

Mark as done

- **More details will be provided in the first exercise session!**

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Schedule (pre-final)

Date	Day	Time	Event	Person
14.10.2021	THU	12:30-14:00	1. Lecture (Upload Ex1)	Decker
15.10.2021	FRI	10:30-12:00	2. Lecture	Decker
21.10.2021	THU	12:30-14:00	1. Exercise (Upload Ex2)	Karim, Welten
22.10.2021	FRI	10:30-12:00	3. Lecture	Decker
28.10.2021	THU	12:30-14:00	4. Lecture	Decker
29.10.2021	FRI		**No Lecture**	
04.11.2021	THU	12:30-14:00	2. Exercise (Upload Ex3)	Karim, Welten
05.11.2021	FRI	10:30-12:00	5. Lecture	Decker
11.11.2021	THU	12:30-14:00	6. Lecture	Decker
12.11.2021	FRI	10:30-12:00	7. Lecture	Decker
18.11.2021	THU	12:30-14:00	3. Exercise(Upload Ex4)	Karim, Welten
19.11.2021	FRI	10:30-12:00	8. Lecture	Decker
25.11.2021	THU	12:30-14:00	9. Lecture	Decker
26.11.2021	FRI	10:30-12:00	10. Lecture	Decker
02.12.2021	THU	12:30-14:00	4. Exercise(Upload Ex5)	Karim, Welten
03.12.2021	FRI	10:30-12:00	11. Lecture	Decker
09.12.2021	THU	12:30-14:00	12. Lecture	Decker
10.12.2021	FRI		**No Lecture**	
16.12.2021	THU		**No Lecture**	
17.12.2021	FRI	10:30-12:00	5. Exercise(Upload Ex6)	Karim, Welten
23.12.2021	THU		**No Lecture**	
06.01.2022	THU	12:30-14:00	13. Lecture	Decker
07.01.2022	FRI	10:30-12:00	14. Lecture	Decker
13.01.2022	THU	12:30-14:00	6. Exercise	Karim, Welten
09.02.2022	Di.	14:00-16:00	Exam 1	Karim, Welten
10.03.2022	Do.	12:00-14:00	Exam 2	Karim, Welten

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY
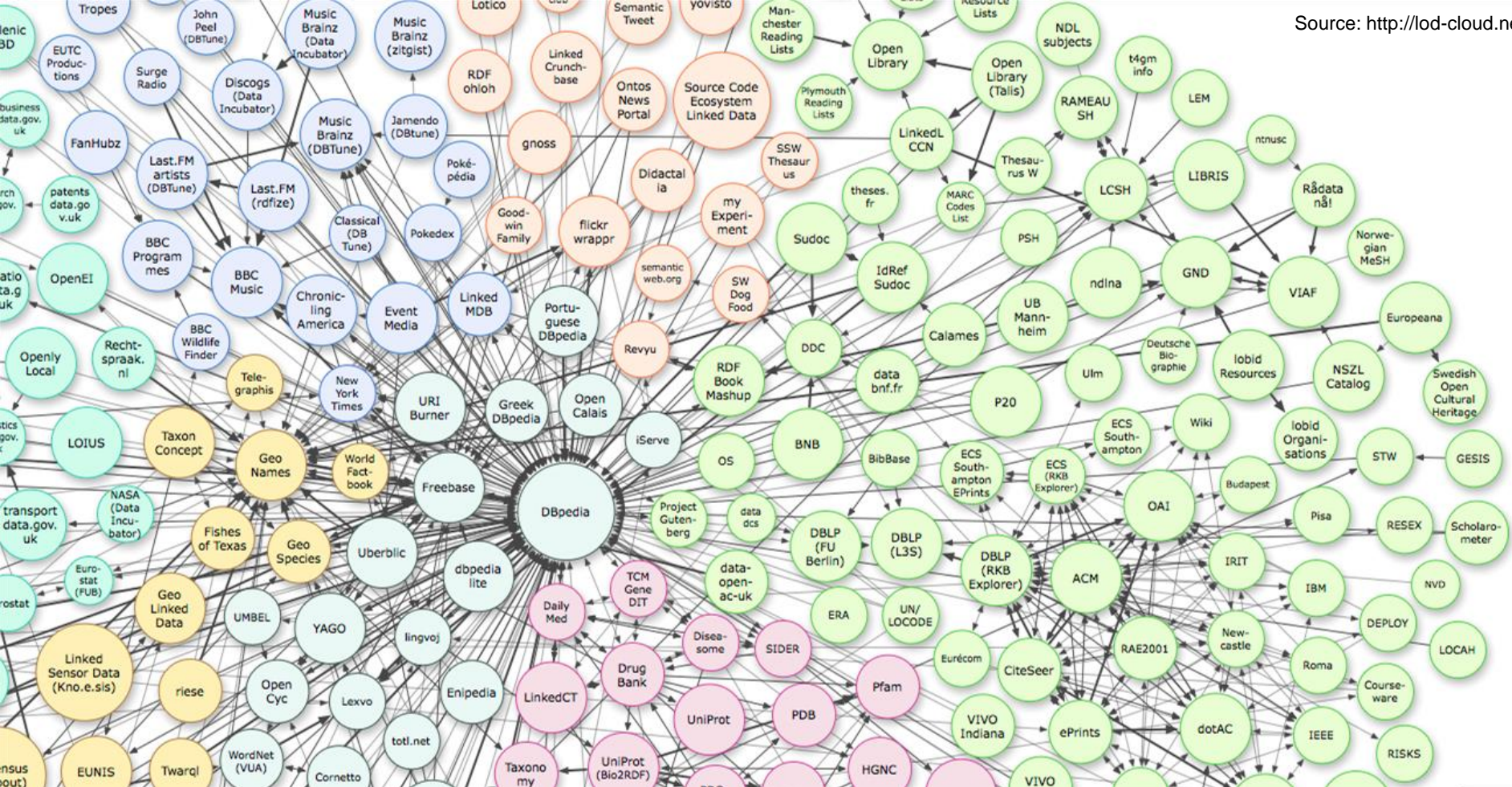
Semantic Web Lecture

C01 Hypertext, the Internet and the Web
How does the web work?

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

CC - Creative Commons Licensing

- This set of slides is part of the lecture „Semantic Web Technologies" held at RWTH Aachen University
- The slides are partially based on slides of Benedikt Köhler and Andreas Harth and were prepared by Lars Gleim

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

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

Semantic Web Lecture
Prof. Stefan Decker
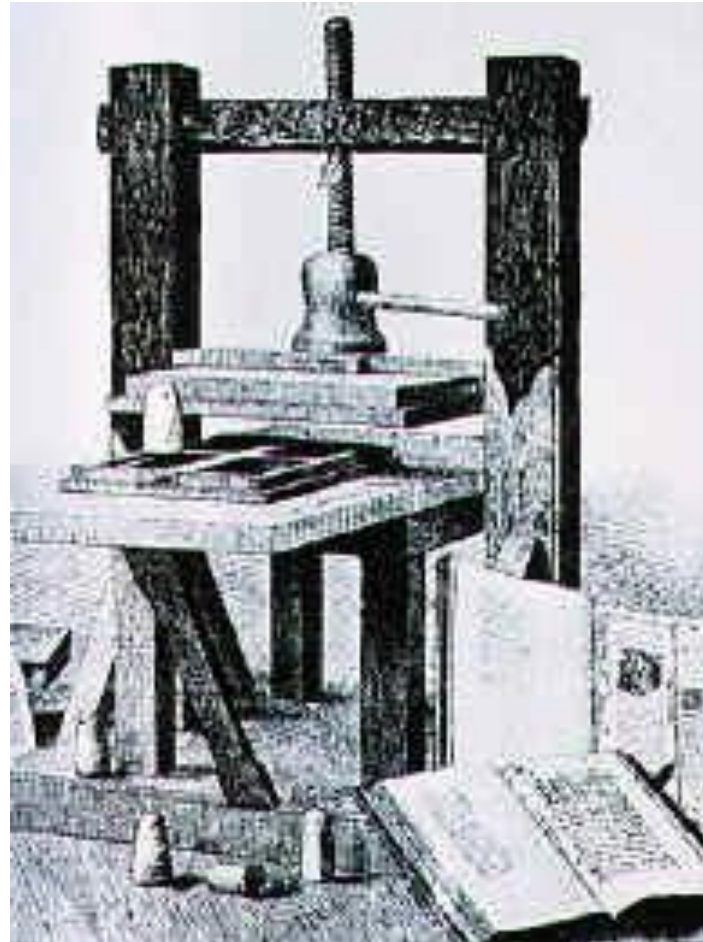Lehrstuhl Informatik 5

**Informatik 5 -
Informationssysteme
und Datenbanken**

RWTH AACHEN UNIVERSITY

Cave Drawings (30000 BC)

Writing (Sumerian cuneiform 3200 BC)

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Printing Press (Gutenberg 1450)

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Photography (Daguerre 1839)

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Phonograph (Edison 1877)

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Movies (Lumiere 1895)

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

14

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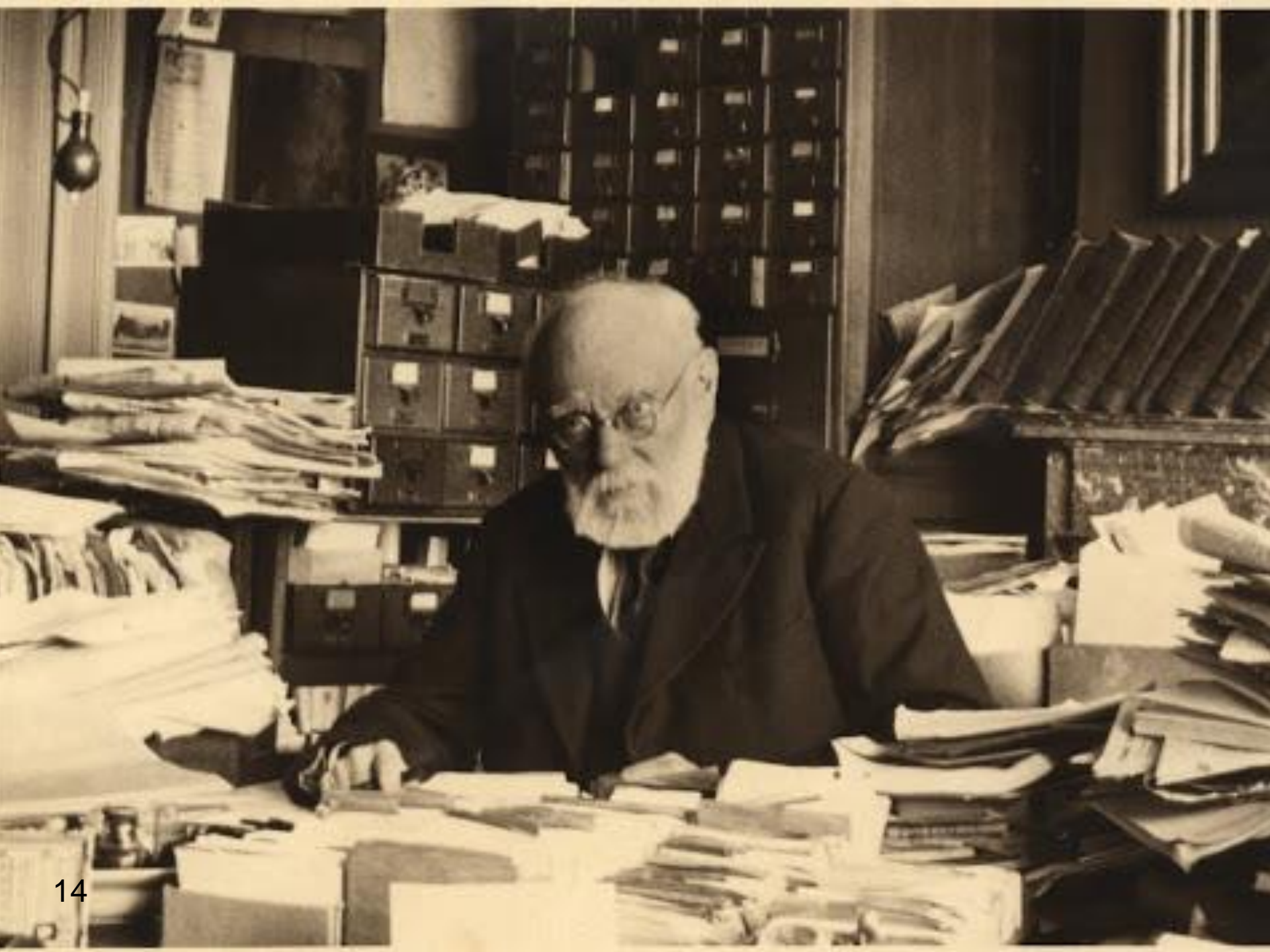
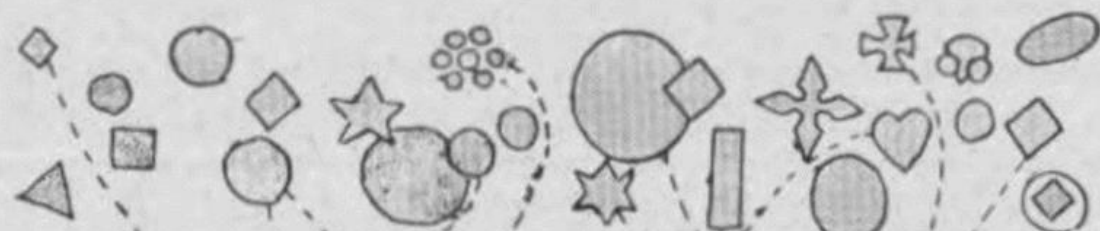|	
(4)	Europe (continent).
(49)	Petits pays de l'Europe (groupement de pays).
(493)	Belgique (pays).
(493.2)	Brabant (province).
(493.21)	Bruxelles (arrondissement).
(493.211)	Bruxelles (canton).
(493.211.1)	Bruxelles (ville).

Les choses

nivers, la Réalité, le Cosmos

es intelligences

sent les choses fragmentairement
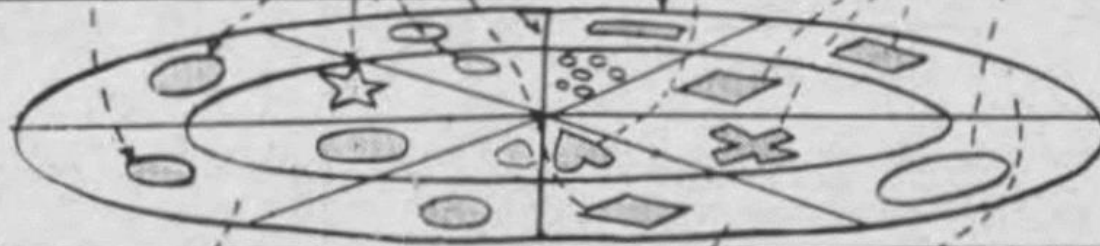
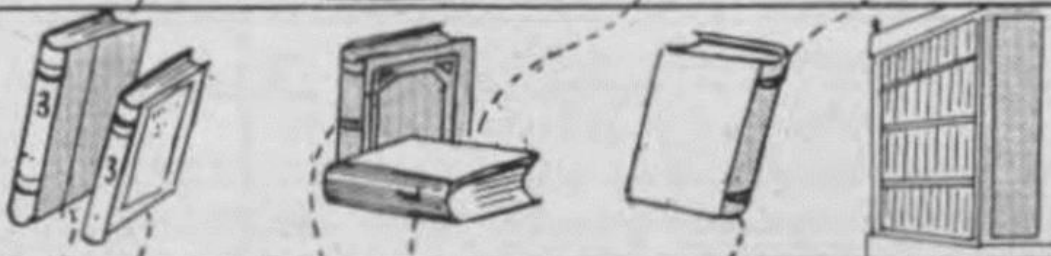## La science

oordonne en ses cadres les pensées
les intelligences particulières

Les Livres

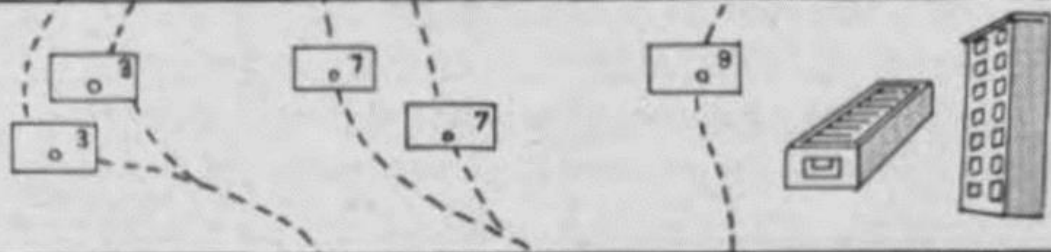vent et photographient la science
dre divisé des connaissances
ion de livres forment la Bibliotheque

La Bibliographie
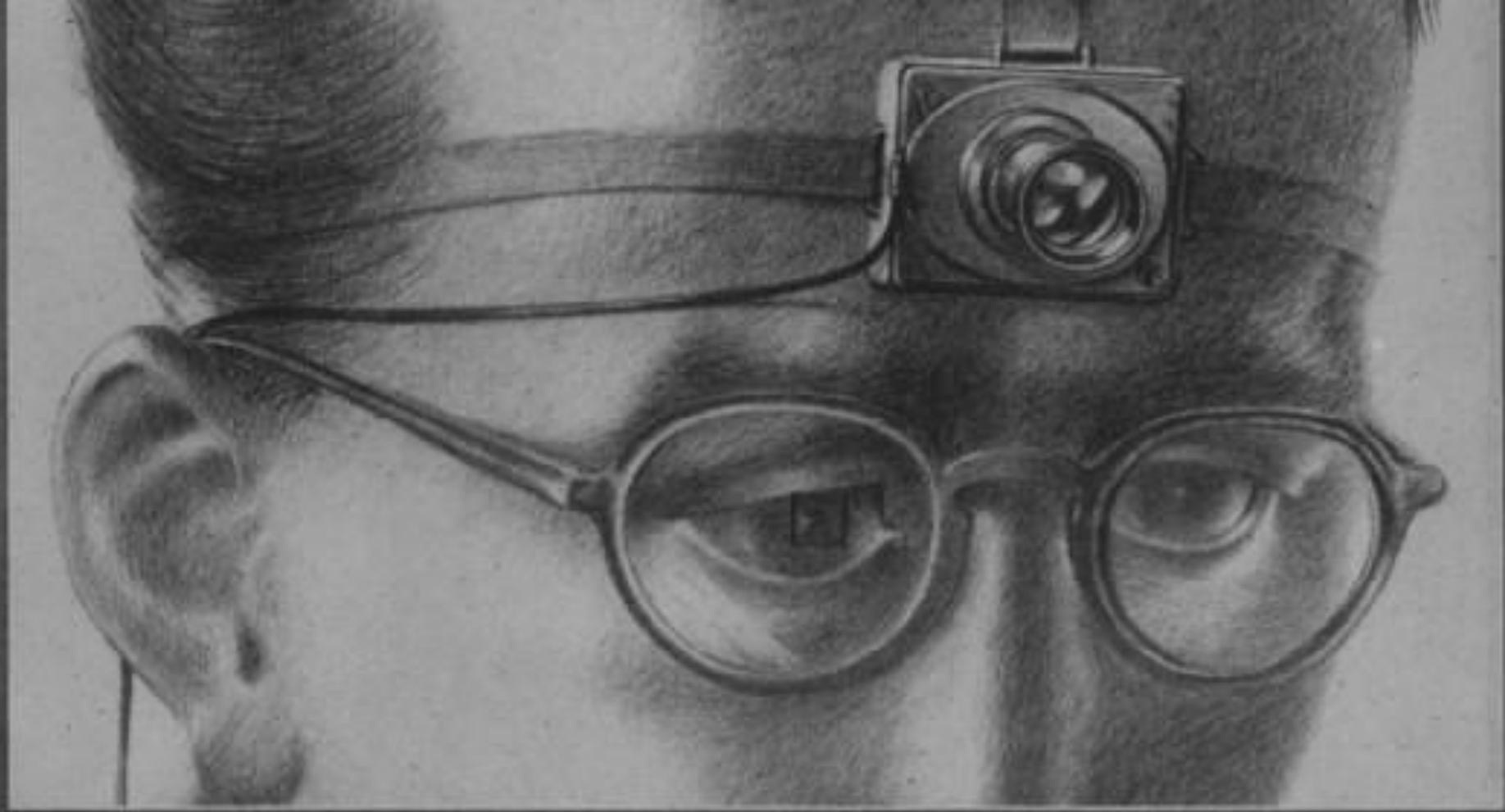
rie et catalogue les Livres

16

n de notices Bibliographiques forme
ertoire Bibliographique universel

A SCIENTIST OF THE FUTURE RECORDS EXPERIMENTS WITH A TINY CAMERA FITTED WITH UNIVERSAL-FOCUS LENS. THE SMALL SQUARE IN THE EYEGLASS AT THE LEFT SIGHTS THE OBJECT

AS WE MAY THINK

A TOP U. S. SCIENTIST FORESEES A POSSIBLE FUTURE WORLD IN WHICH MAN-MADE MACHINES WILL START TO THINK
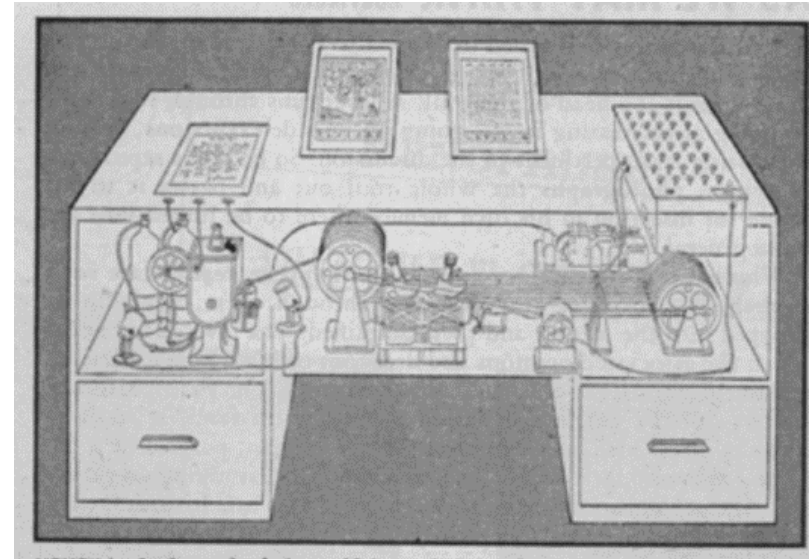
The Memex (1945)
Vannevar Bush in "As We May Think"
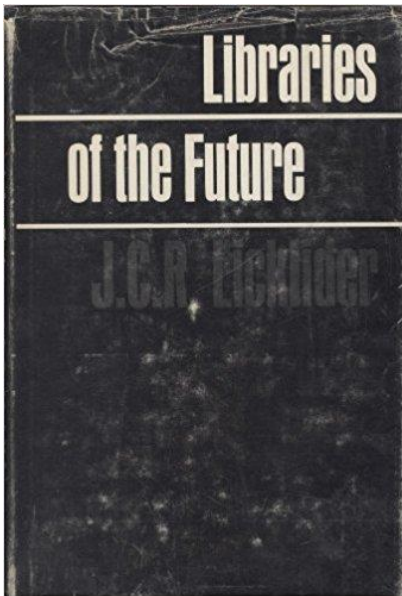## *The Atlantic Monthly, July 1945*

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 . . .

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Libraries of the Future…1965

"It should be possible, for example, to transfer an entire system of chemical formulas directly from the general fund of knowledge to a chemical process-control system, and to do so under human monitorship but not through human reading and key pressing."

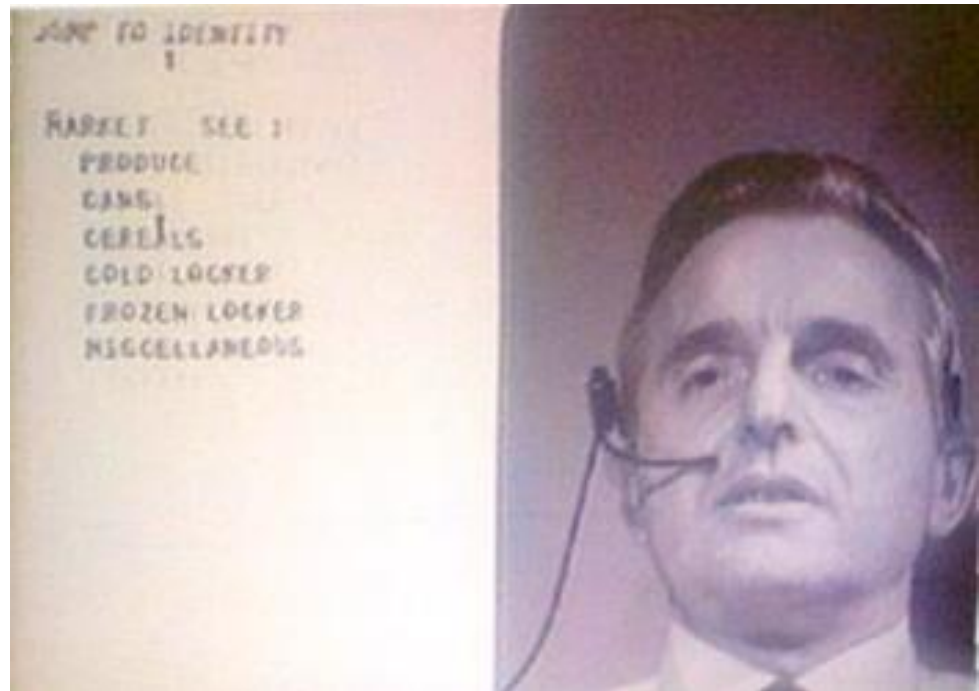J. C. R. Licklider. Libraries of the Future. The MIT Press. 1965.

oNLine System- NLS (Doug Engelbart, SRI, 1968)

"By 'augmenting human intellect' we mean increasing the capability of a man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems." AUGMENTING HUMAN INTELLECT:

A CONCEPTUAL FRAMEWORK, Douglas C. Engelbart, October 1962

- The Mouse;
- Word Processing;
- Data Sharing;
- Hypertext;

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

ARPANET (John Postel, David Crocker, Vint Cerf, 1969)
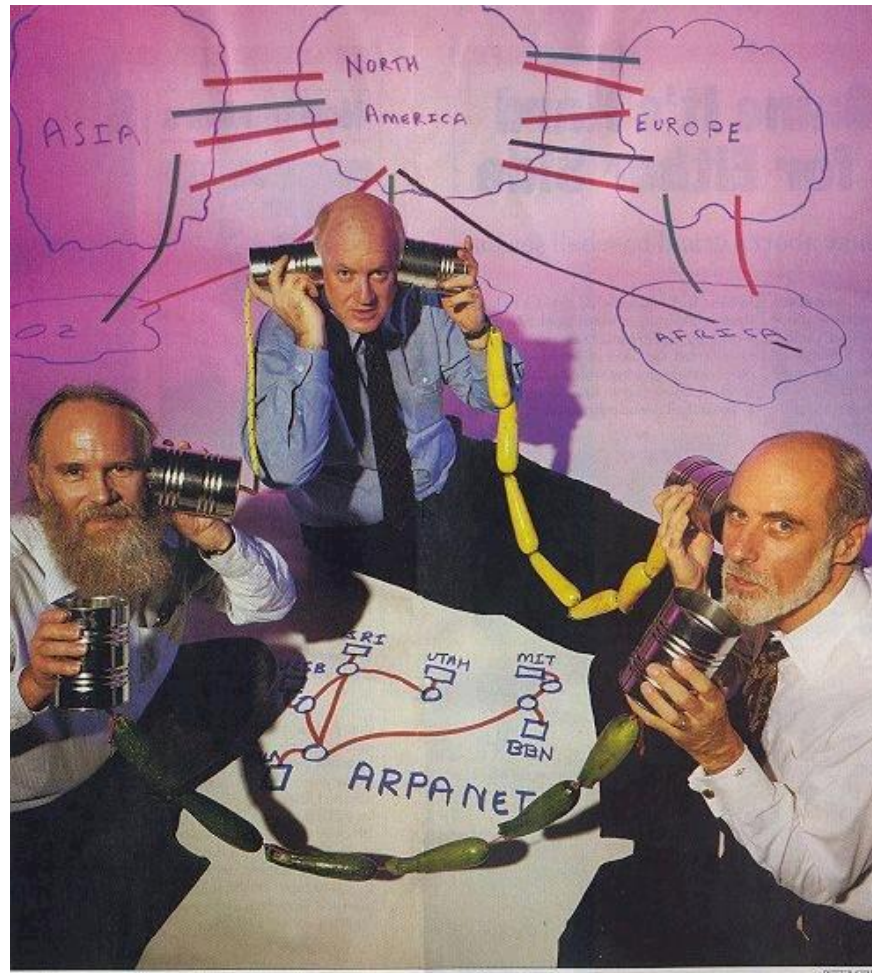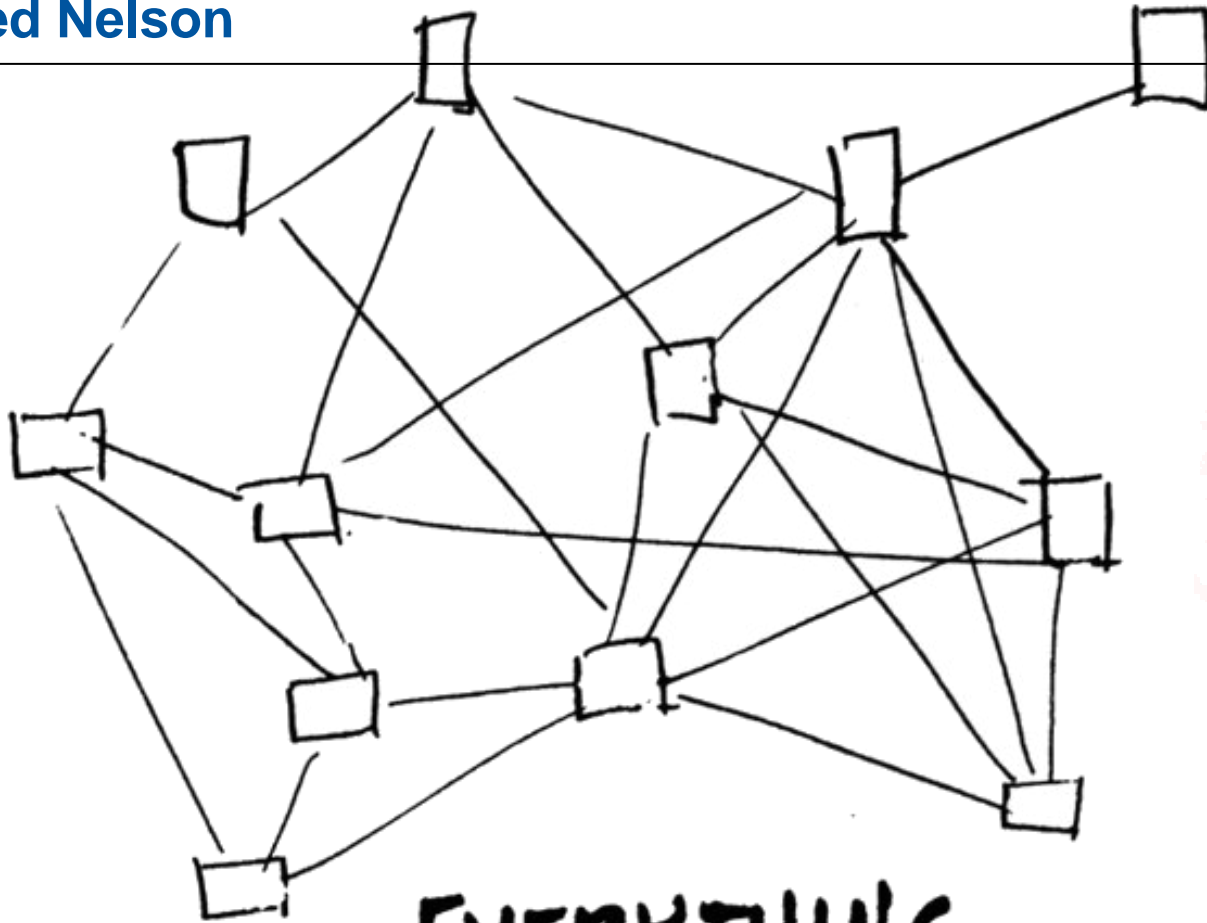
Ted Nelson

EVERYTHING IS DEEPLY INTERTWINGLED.

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Xanadu (Ted Nelson ~1960-???)

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

World Wide Web (Tim Berners-Lee 1989)
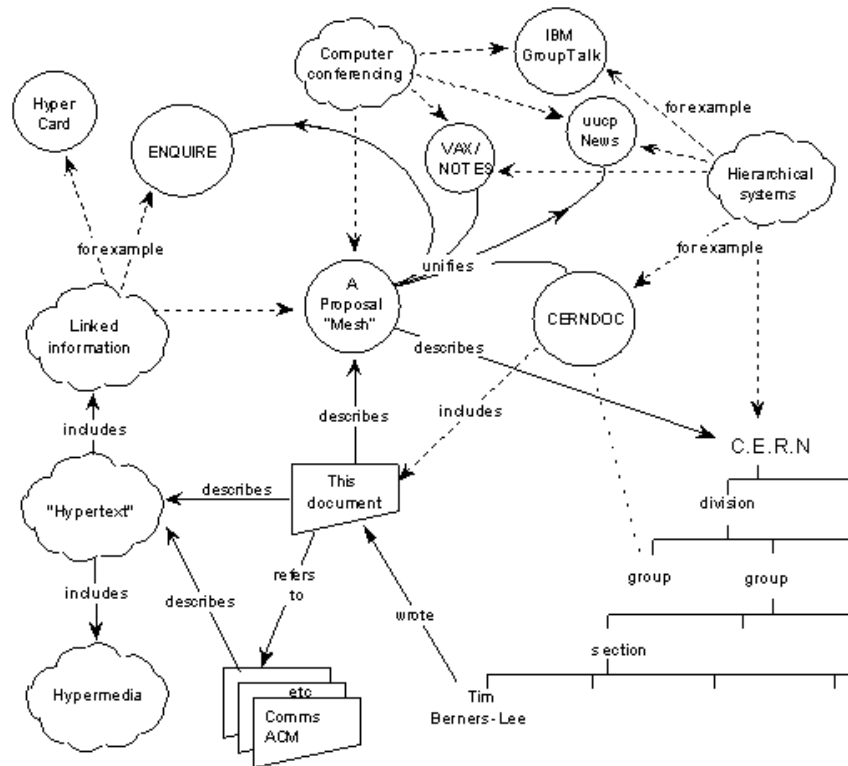
"There was a second part of the dream […] we could then use computers to help us analyse it, make sense of what we re doing, where we individually fit in, and how we can better work together."

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

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

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

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

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Internet Engineering Task Force

- The Internet Engineering Task Force (IETF) is hosted at an organisation called the Internet Society. [1]

- 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/

Development Through RFCs

- RFCs are an example of how to manage standardization in fast developing surroundings

- Proposed specifications have certain status categories[1] (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 obtained[2]

[1] http://tools.ietf.org/html/bcp9#section-4
[2] http://tools.ietf.org/html/bcp9#section-4.1.3

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Postel's Robustness Principle

[1]

- Jon Postel (1943-1998): former RFC editor[2]
  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]"

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

---
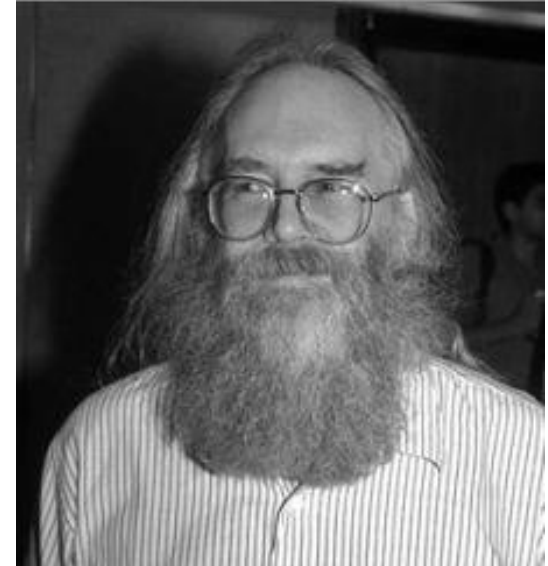
[1] http://www.internethalloffame.org/inductees/jon-postel
[2] Posthumously, he was honored with his own RFC where he is described as leader, icon, giant and „Internet Assigned Numbers Authority" (https://tools.ietf.org/html/rfc2468)
[3] https://tools.ietf.org/html/rfc793#section-2.10

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

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

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
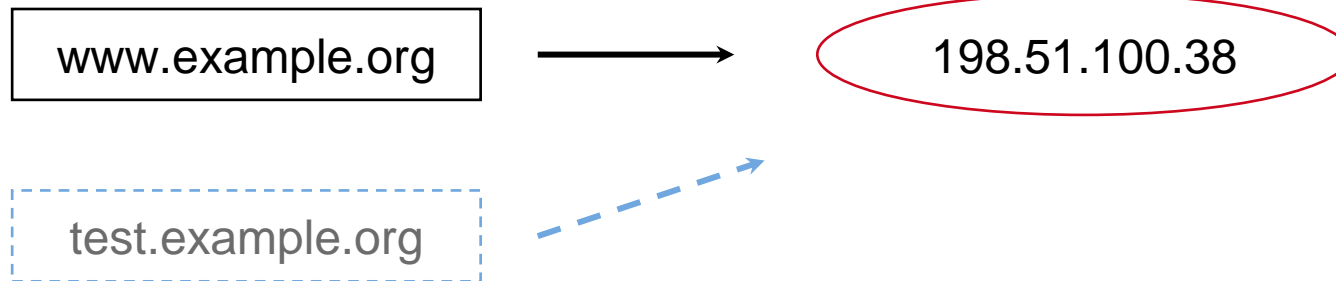und Datenbanken

RWTH AACHEN UNIVERSITY

Internet Assigned Numbers Authority

- The Internet Assigned Numbers Authority (IANA) is a department of ICANN that [1]
    - 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

```
┌─────────────────────┐                    ⬭──────────────────⬮
│   www.example.org   │  ──────────▶        │   198.51.100.38   │
└─────────────────────┘                    ⬮──────────────────⬭

┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐
    test.example.org        ─ ─ ─ ─ ▶
└ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘
```

Multiple hostnames can map to the same IP!

[1] http://iana.org/

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

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)
  Introduction to the Domain Name System (DNS)

4

[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

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
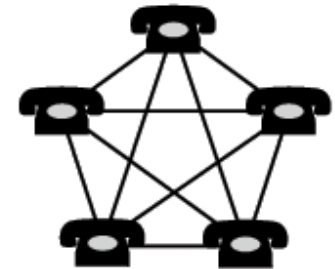Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY
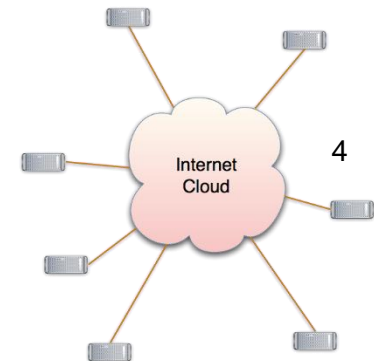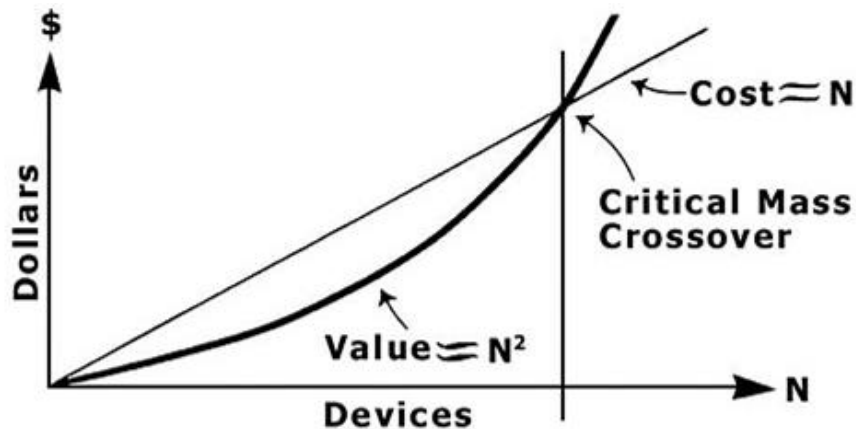
The Network Effect aka Metcalfe's Law

"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 N2 ."

Metcalfe's Law

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY
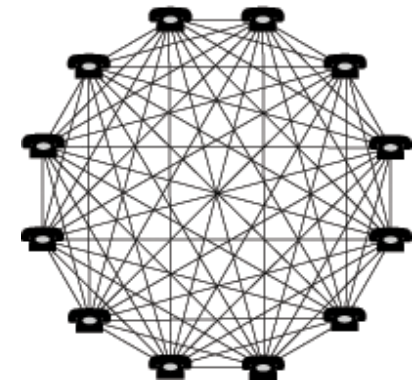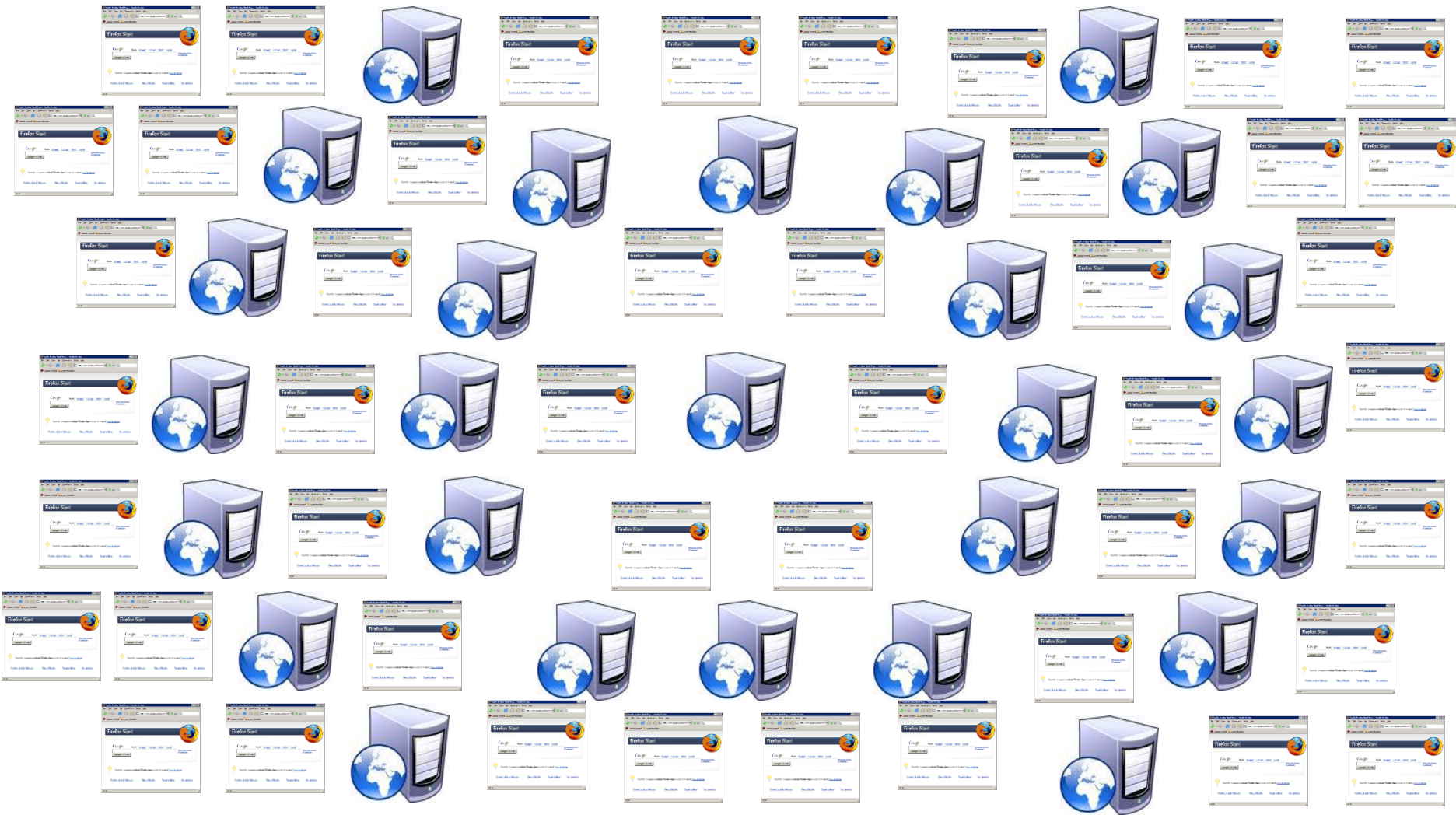
Agenda

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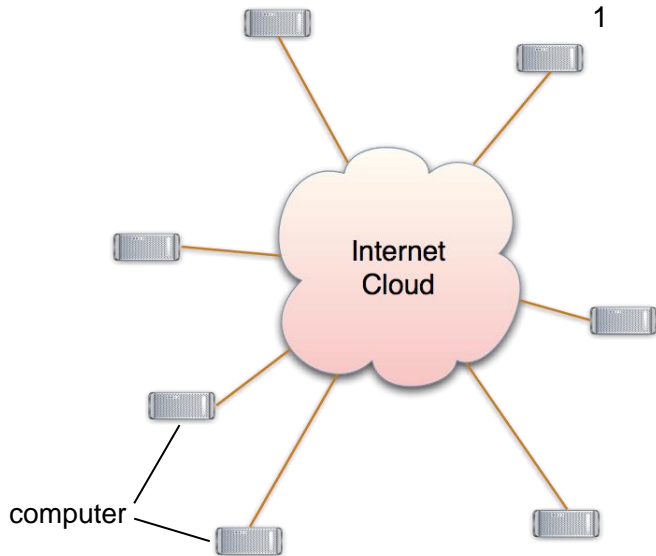   5. User Agents and Servers
   6. The World Wide Web Consortium
4. Towards a Web of Data

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

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
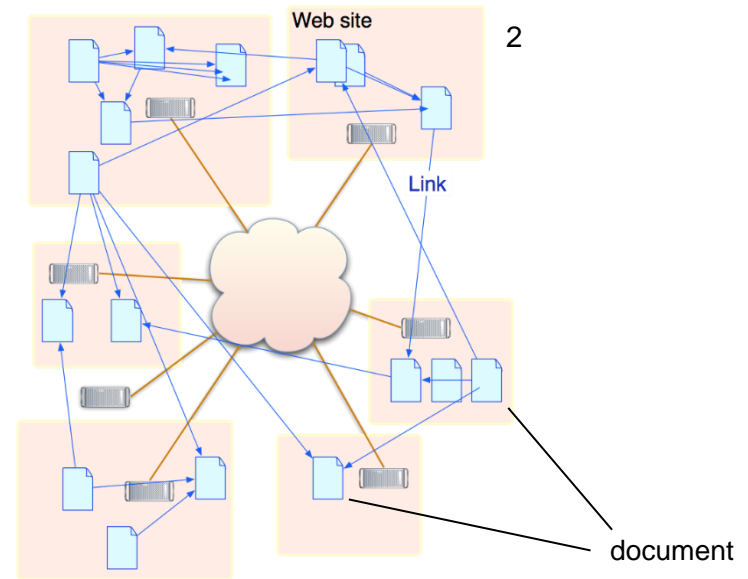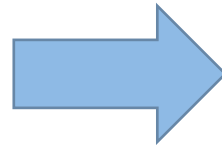
[1,2] http://www.w3.org/DesignIssues/Abstractions.html (great wrap-up of the developing understanding of the internet & web)

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

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

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Resources

Definition 1 (Resource) *A resource is an abstract notion for things of discourse, be they abstract or concrete, physical or virtual.*

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"

Uniform Resource Identifiers

Definition 2 (Uniform Resource Identifier) *A Uniform Resource Identifier (URI) is a character sequence that identifies a resource.*

- I. e. URIs are names for resources
- URIs are defined in RFC 3986[1] (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 URIs[1]

- 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

Structure of URIs

- The generic URI syntax consists of the scheme, authority, path, query, and fragment

- URI syntax specification in RFC 3986:

```
    URI  = scheme ":" hierarchical part[ "?" query ][ "#"
fragment ]
```

- Within `[  ]`: optional parts

- `hierarchical part`[1]: contains the authority, headed by "//" and followed by an optional path (tracking the hierarchical server structure leading to the resource)

[1] This is only the most common setup of the hierarchical part, used majorly on the internet and sufficient for our purposes. Some less frequently used URIs follow another syntax. For an complete definition of the URI syntax, see RFC 3986 p.15 et seq.

Structure of URIs

- An example of a HTTP URI looks like the following:

```
http://example.com:8042/over/there?name=ferret#nose
```

scheme   authority   path   query   fragment
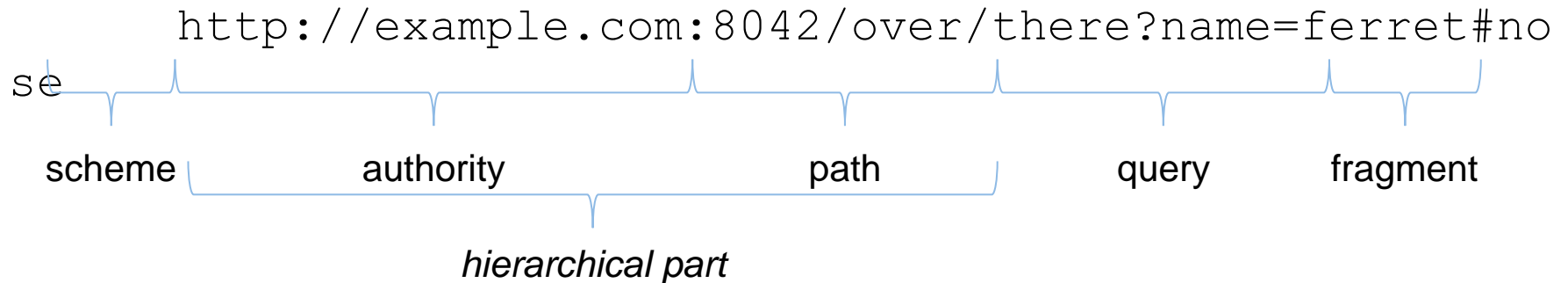
hierarchical part

- **scheme**: specifies type of URI (and the associated protocol1)
- **authority**: hostname and optional port number
- **path**: reference to content on the host server; can be interpreted in a hierarchical fashion (similar to a directory tree on a file system)
- **query**: contains data that does not match tree structure (variables)
- **fragment**: addresses local part of a document

[1] View IANAs registered schemes: http://www.iana.org/assignments/uri-schemes/uri-schemes.xhtml

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

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 of the World Wide Web) is one of them

FTP
File Transfer Protocol

SMTP
Simple Mail Transfer Protocol

Internet

and many others!

HTTP
Hypertext Transfer Protocol

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY
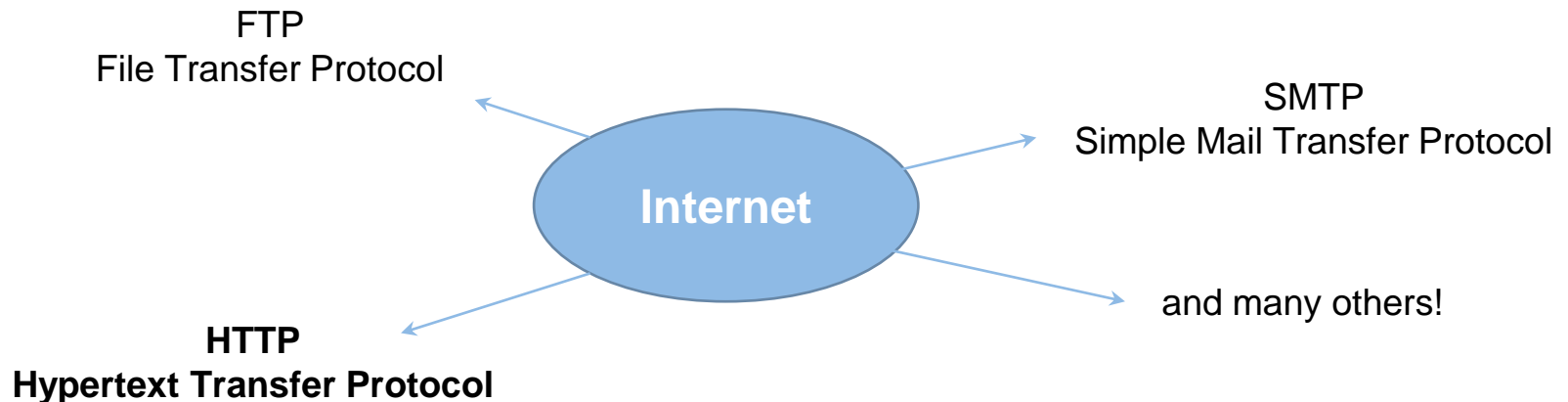
Hypertext Transfer Protocol

- On the web, the communication between a user agent and a server is carried out via the Hypertext Transfer Protocol (HTTP)

HTTP 1.1 is defined in a set of RFCs (all „Proposed Standard"[1])	
RFC 7230	**Message Syntax and Routing:** Basic structure of request and response, including basic header fields
RFC 7231	**Semantics and Content:** Request methods, status codes and additional header fields
RFC 7232	**Conditional Requests**
RFC 7233	**Range Requests**
RFC 7234	**Caching**
RFC 7235	**Authentication**

- The ones of interest to us are RFC 7230 and RFC 7231

[1] http://www.rfc-editor.org/categories/rfc-proposed.html

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTHAACHEN
UNIVERSITY

Hypertext Transfer Protocol - Definition

- RFC 7230[1] (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, but also household appliances, cars, internet radio devices or command line programs

- Servers can be: traditional large-scale web servers of commercial companies, but also devices that offer access to data or functionality, (household appliances, cars,...)

[1] http://tools.ietf.org/html/rfc7230

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Successful Request/Response Interaction

- A simple successful request/response interaction between user agent and server works as follows:

    - The user agent establishes a TCP connection with the server.
    - The user agent transmits a HTTP request. Trailing newline to indicate end of message.
    - The server replies with a HTTP response. Trailing newline to indicate end of message.
    - The connection is terminated.

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

UML Sequence Diagram
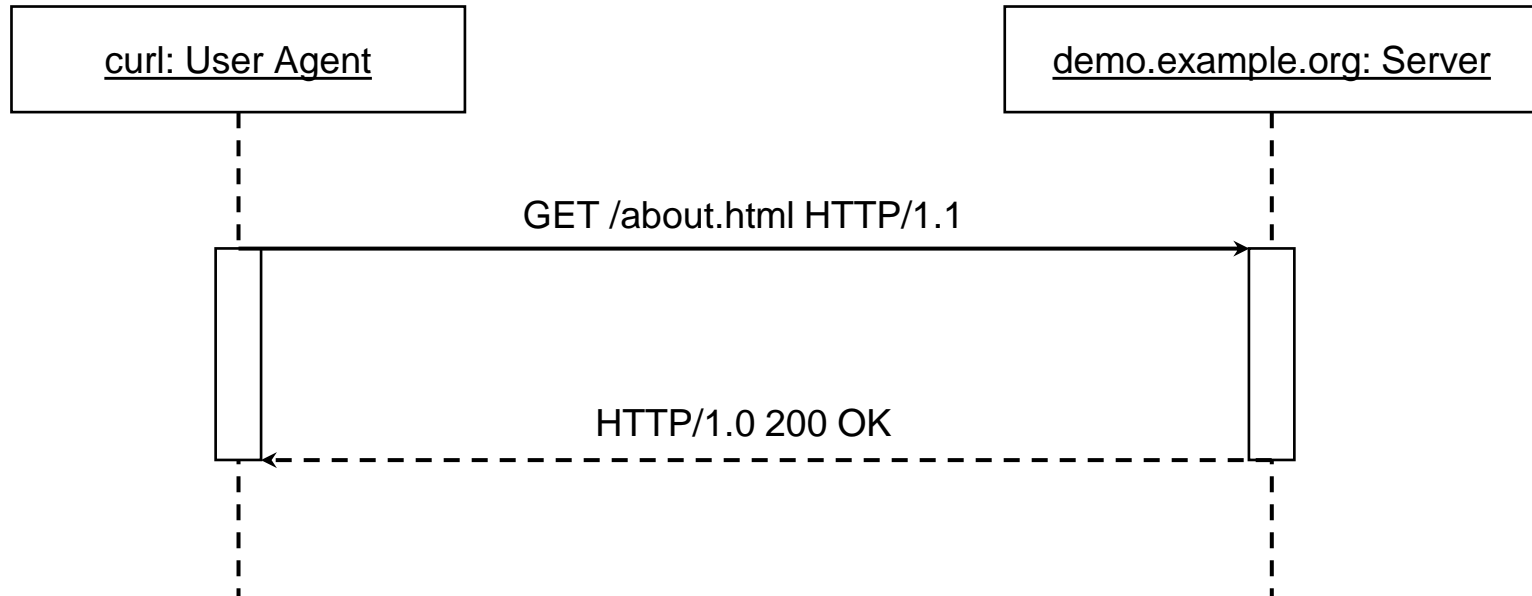
HTTP – Stateless Protocol

- HTTP is a stateless protocol and every transaction is treated independently by the server[1]

- Even if on the same connection, the server must not assume that two different requests are issued by the same user agent

- In many cases on the web, there is a need for permanent authentications: staying logged-in, shopping baskets, …

- This can be achieved through different means of session management:
  - HTTP authentication
  - Session-IDs
  - Cookies

[1] http://tools.ietf.org/html/rfc7230#page-11

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

HTTP Request/Response - Structure

- Both HTTP request and response messages follow the same structure:

HTTP request:

HTTP response:

1 Start line:
The request line for HTTP requests, status line for HTTP responses. The start line is mandatory.

2 Header fields:
Additional information about the message in key/value format. Header fields are optional.

3 Message body:
The transferred content (a list of bytes). Message body is optional.

1 `GET /about.html HTTP/1.1`

2 `Host: demo.example.org`
`User Agent: curl/7.2.10`
`Accept */*`

3

1 `HTTP/1.0 200 OK`

2 `Server: Apache`
`Content-Type: text/html`

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

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

HTTP Methods

- The HTTP method is part of the request line

- Via the HTTP methods, the user agent informs the server about the purpose of the request

- There are several registered HTTP methods[1], we only focus on the GET method (see RFC 7231[2])

[1] http://www.iana.org/assignments/http-methods/http-methods.xhtml
[2] https://tools.ietf.org/html/rfc7231#section-4.1

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

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

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Selected HTTP Status Codes

Status code	Description
200 OK	The request to the target resource succeeded.
300 Multiple Choices	Indicates multiple content options for the user agent to choose from.
301 Moved Permanently	All requests should be directed to another URI.
303 See Other	The response to the request can be found under the given URI.
401 Unauthorized	Authentication is required and has not yet been given.
404 Not Found	The requested resource cannot be found.
410 Gone	The resource has been deleted.
500 Internal Server Error	Generic error messages indicating problems on the server.

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Selected HTTP Header Fields

Header field	Used by*	Description
Host	U	The TCP connection is to an IP address. One server might host multiple web sites with different hostnames. With Host, the user agent specifies the hostname. Host is required in HTTP/1.1.
User-Agent	U	Name of user agent
Server	S	Name of the server
Accept	U	The agent can specify acceptable media types
Content-Type	U, S	The media type of the message body
Location	S	The location of the requested URI, for example in case of redirects (status codes 3xx)
Content-Location	S	The content of the message body is really available at the specified location, not the requested URI

* U = user agent header, used in the request; S = server header, used in response.

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
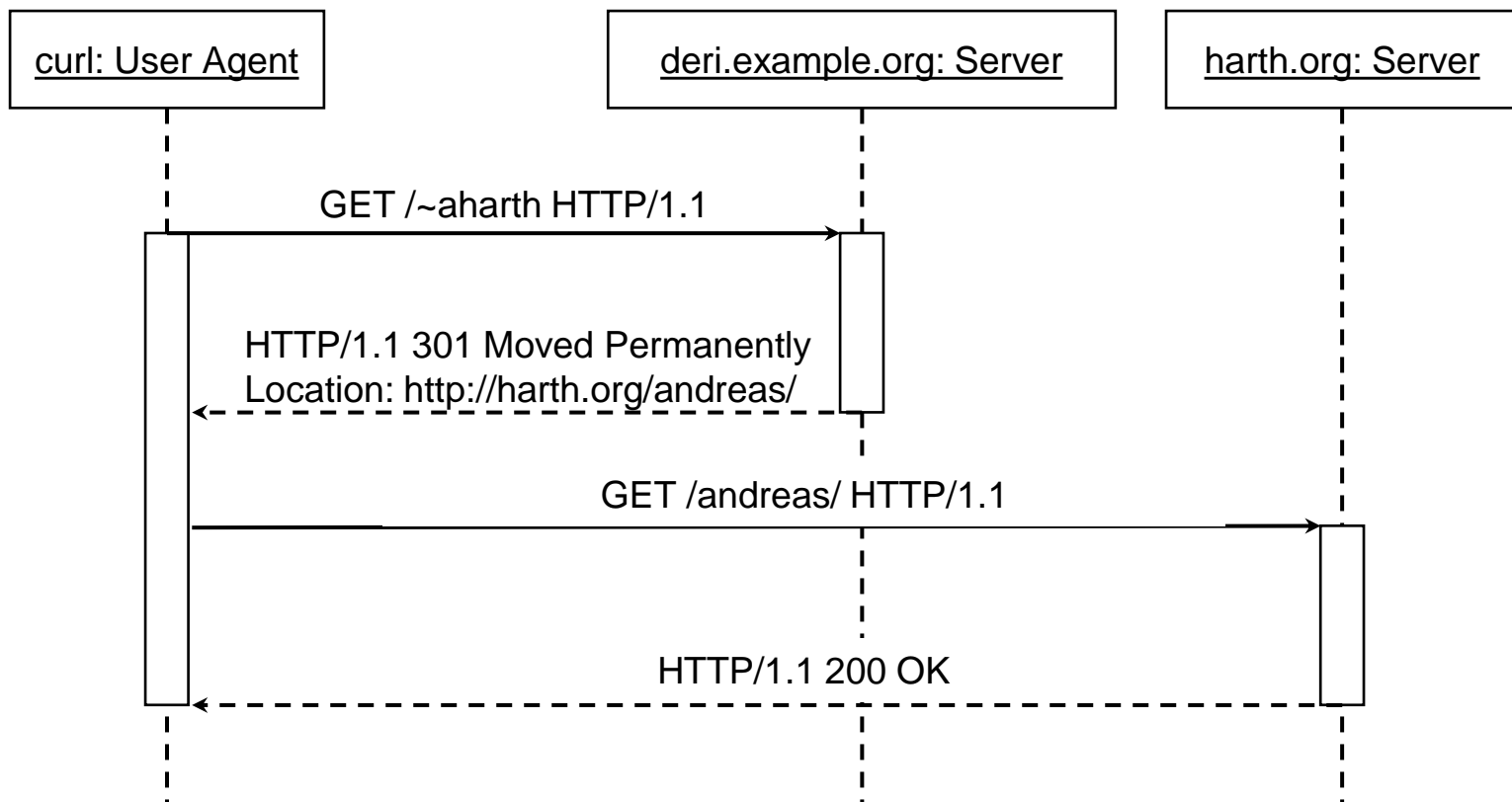UNIVERSITY

HTTP Message Body

- The message body contains a byte stream of content (e.g., HTML document, PNG picture)

- In principle, only the server response contains a message body (the requested content)

- However, there are cases where the request includes a message body, for example when uploading a file from the browser

- To know how parse the bytes contained in the message body, the recipient needs information about the type of content

- The process of figuring out the media type of messages is called content negotiation

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

HTTP Message, Request, Response

Definition 4 (HTTP Message, HTTP Request, HTTP Response) *A HTTP message is a tuple $\langle S, H, B \rangle$, where S is the mandatory start line, H is an optional list of header name/value pairs, and B is the message body, also optional. A HTTP request is a HTTP message, in which the start line S consists of a tuple $\langle M, u, V \rangle$, where M is the HTTP method, $u \in \mathcal{U}$ is the request URI and V is the HTTP version. A HTTP response is a HTTP message, in which the start line S consists of a tuple $\langle SC, V \rangle$, where SC is the HTTP status code and V is the HTTP version.*

HTTP Interaction

Definition 5 (HTTP Interaction) *The function* http *models the execution of a HTTP request that leads to a HTTP response. Let req be a HTTP request and resp be a HTTP response. We can write* $resp = \text{http}(req)$.

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

HTTP Redirects with Location Header

- Example of HTTP interaction with a 301 Moved Permanently redirect from http://deri.example.org/~aharth/ to http://harth.org/andreas/.

- Redirect also occur as part of content negotiation

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Content Negotiation[1]

- Typically, the user-agent has a preference about the content-type of the requested data (format, language, encodings,…)

- This preference is indicated in the accept header of the HTTP request

- The server replies with the specified location of the requested file via the Content-Location header. The user runs another HTTP request, now aiming directly at the specified location.

[1] http://tools.ietf.org/html/rfc7231#section-3.4

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Selected Content/Media Types[1]

Media type	Format of message body
`*/*`	Any (makes only sense in `Accept` header)
`text/html`	HTML
`application/xhtml+xml`	XHTML
`application/rdf+xml`	RDF/XML
`application/ld+json`	JSON-LD
`text/turtle`	RDF Turtle
`text/n3`	Notation3

[1] https://www.iana.org/assignments/media-types/media-types.xml

Hypertext Markup Language HTML

- HTML can be used to encode Hypertext documents

- In 1995, HTML 2.0 was specified as an IETF RFC 1866[1]

- The next version, HTML 3.2[2], 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 finalized within the W3C process[3]

[1] http://tools.ietf.org/html/rfc1866
[2] http://www.w3.org/TR/REC-html32/
[3] http://www.w3.org/TR/html5/

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

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!

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Hypertext Markup Language HTML

- For the purpose of our course 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

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY
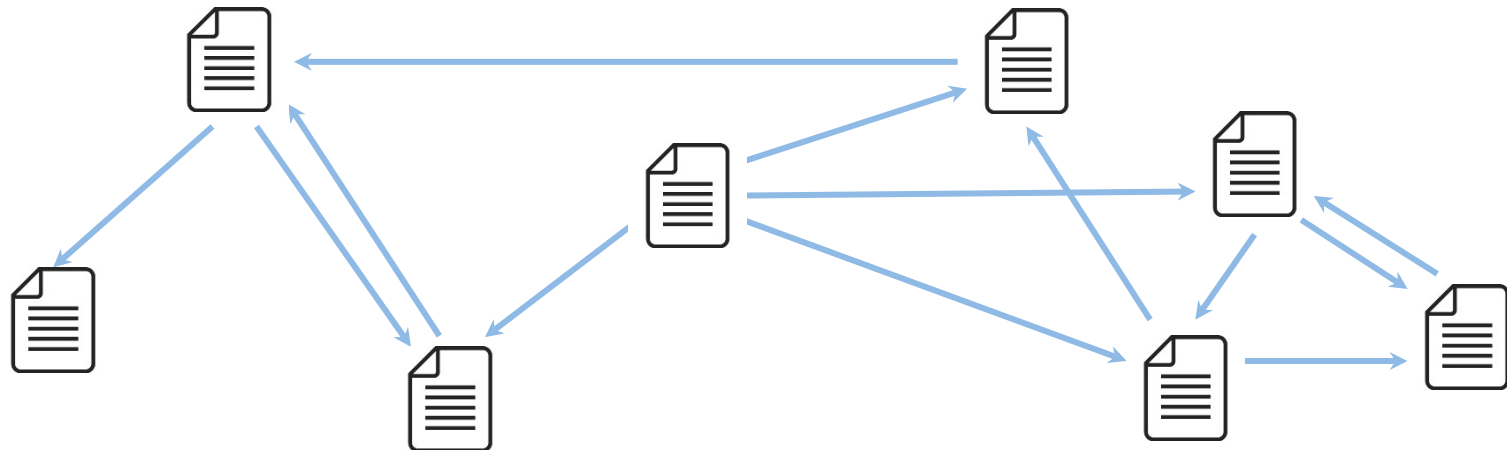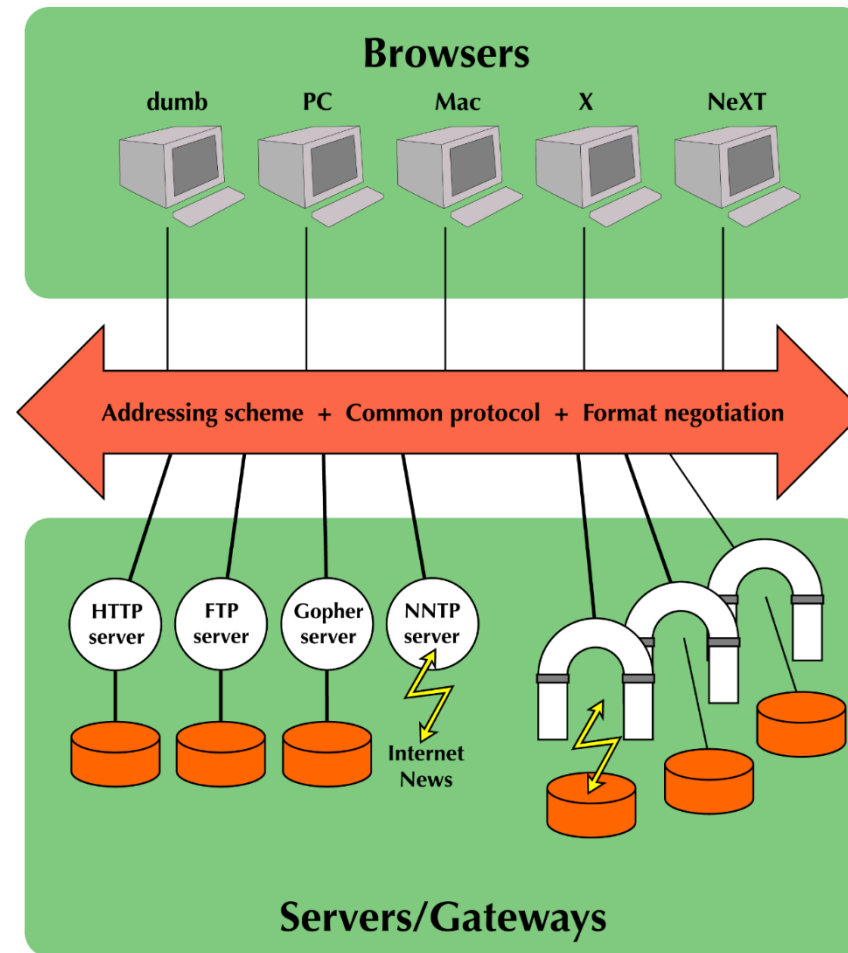
Extensible Markup Language XML

- XML is a format that is readable to both humans and machines

- Thus, 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 values[1]

- XML 1.0 became W3C recommendation in 1998[2]

[1] http://doi.org/10.1109/4236.877487
[2] http://www.w3.org/TR/1998/REC-xml-19980210

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

User Agents and Servers

World Wide Web Consortium W3C

"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]

 [1]

- W3C aims for the worldwide availability of web access and envisions web progress in terms of interaction, data, services, and security through standardization[3]

- 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

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

W3C Drafts and Recommendations

- Comparable to IETF's standardisation through RFCs, W3C publishes reports that pass different maturity levels until they are officially recommended[1]:

  - Working Draft (WD)

  - Candidate Recommendation (CR)

  - Proposed Recommendation (PR)

  - W3C Recommendation (REC)

- In this lecture, we are concerned with languages specified in W3C recommendations[2]

[1] http://www.w3.org/2005/10/Process-20051014/tr.html#maturity-levels
[2] http://www.w3.org/TR/

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Agenda

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

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Making Progress…

Memex (Vannevar Bush)

A memex is "a device in which an individual stores all his books, records, and communications."

Augmenting Human Intellect (Doug Engelbart)

"By "augmenting human intellect" we mean increasing the capability of a man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems."
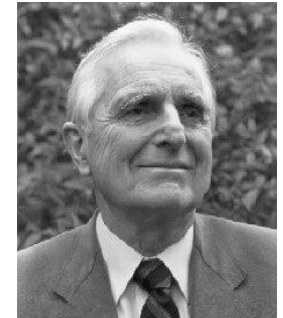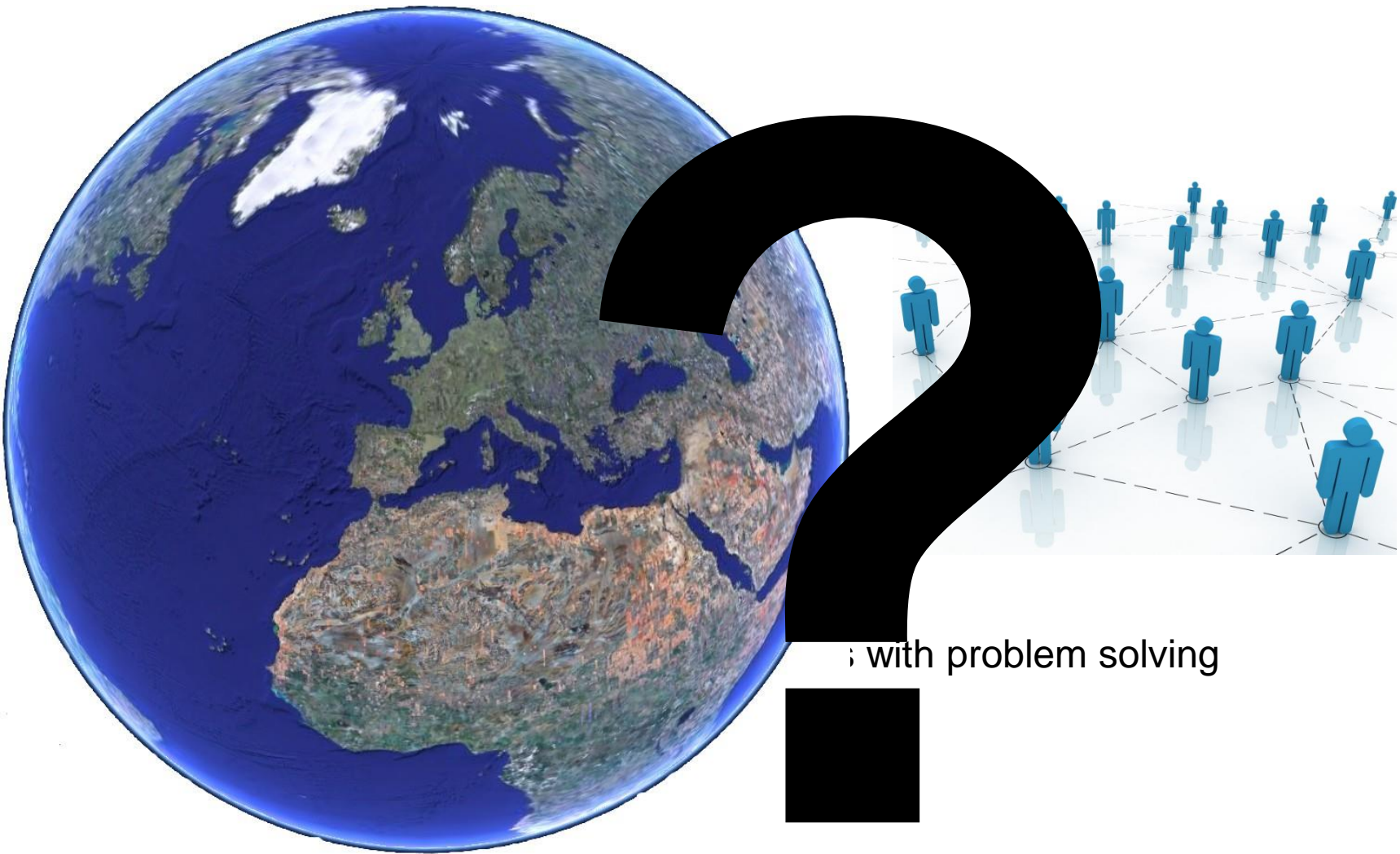
WWW (Tim Berners-Lee)

"There was a second part of the dream […] we could then use computers to help us analyse it, make sense of what we re doing, where we individually fit in, and how we can better work together."

with problem solving

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

What enabled the Web?

Scalability

No growth scalability problem (e.g., no back links from HTML pages)
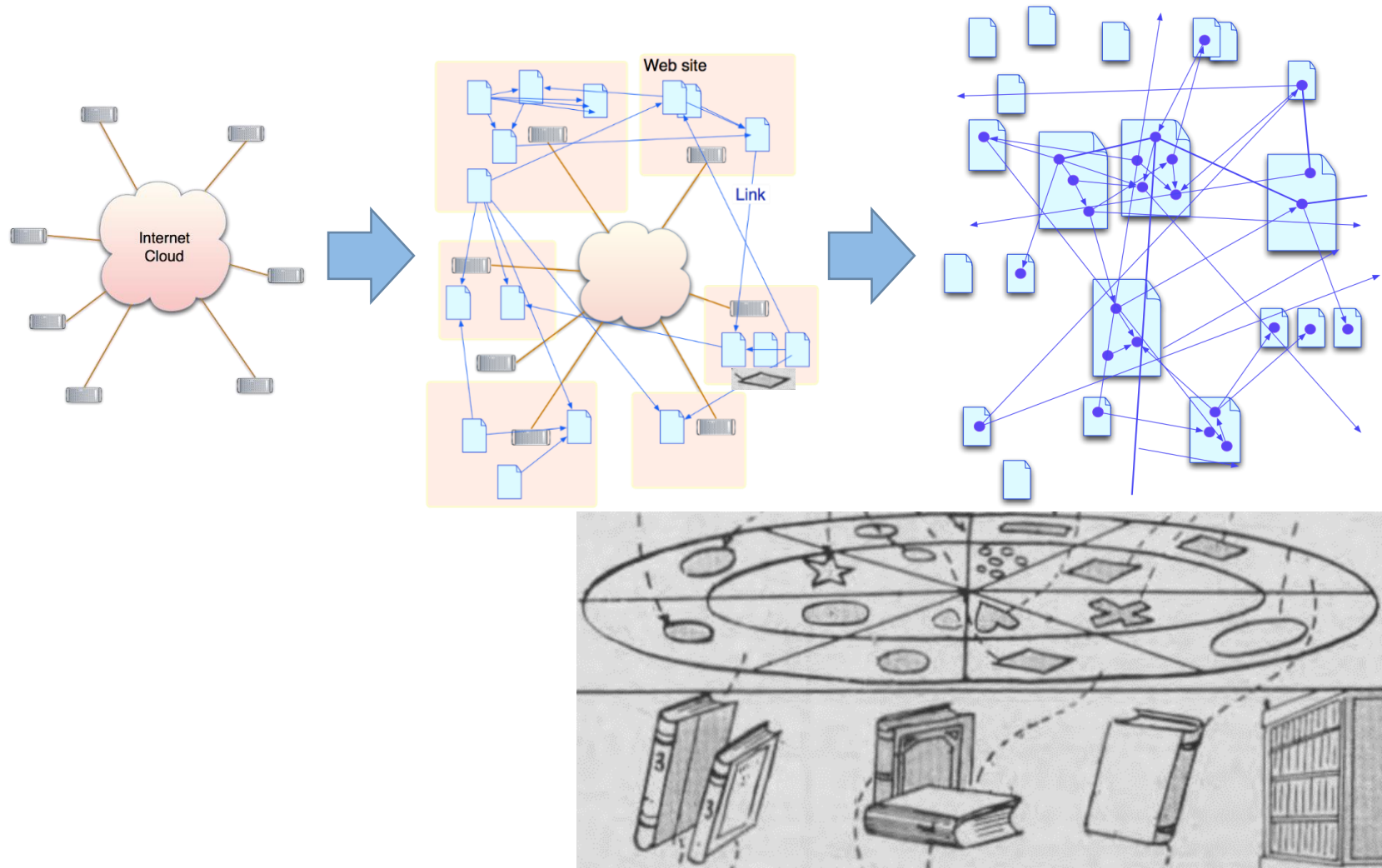
No censorship

No lengthy permission or review process ("permissionless innovation")

Positive feedback loop

Value of network grows with number of participants

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Towards a Web of Data

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Important Distinctions

- We need to distinguish between other resources (Otlet's „frames of thought) and information resources (Otlet's books)
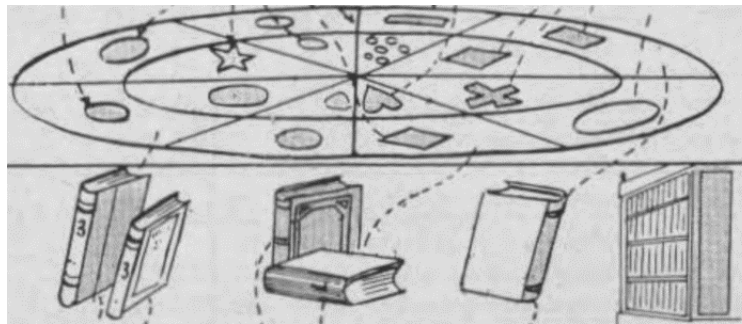
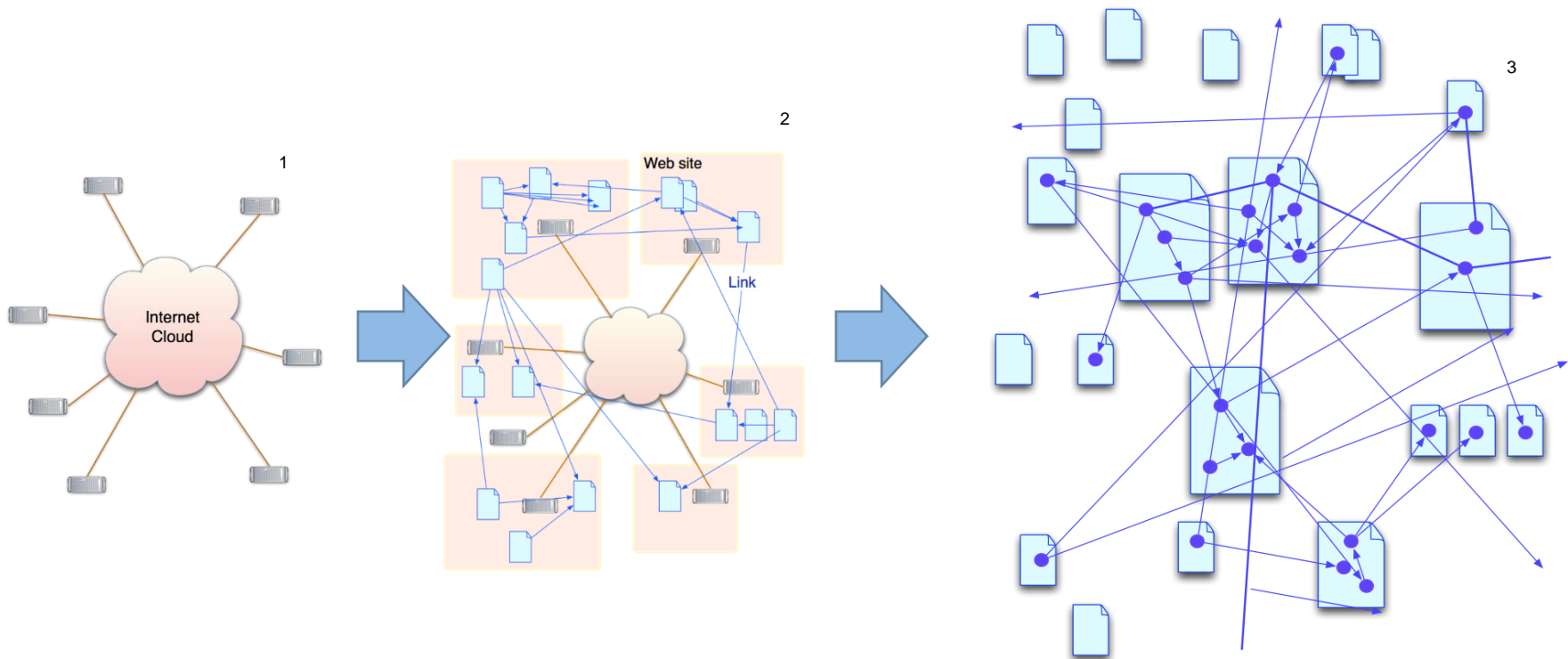- We also need to distinguish between the name of a thing (the URI) and the thing itself (the resource)

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Learning Goals

- G 1.1 Distinguish between hierarchical structure and network structure in representing information.

- G 1.2 Outline basic internet and web technologies, and describe the relation between them.

- G 1.3 Assess the characteristics of decentralisation in information system architectures.

- G 1.4 Explain the network effect as expressed in Metcalfe's Law.

- G 1.5 Identify and name resources; tell information resources apart from other resources based on a textual description.

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Outlook – Chapter 2

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

[1,2,3] http://www.w3.org/DesignIssues/Abstractions.html

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

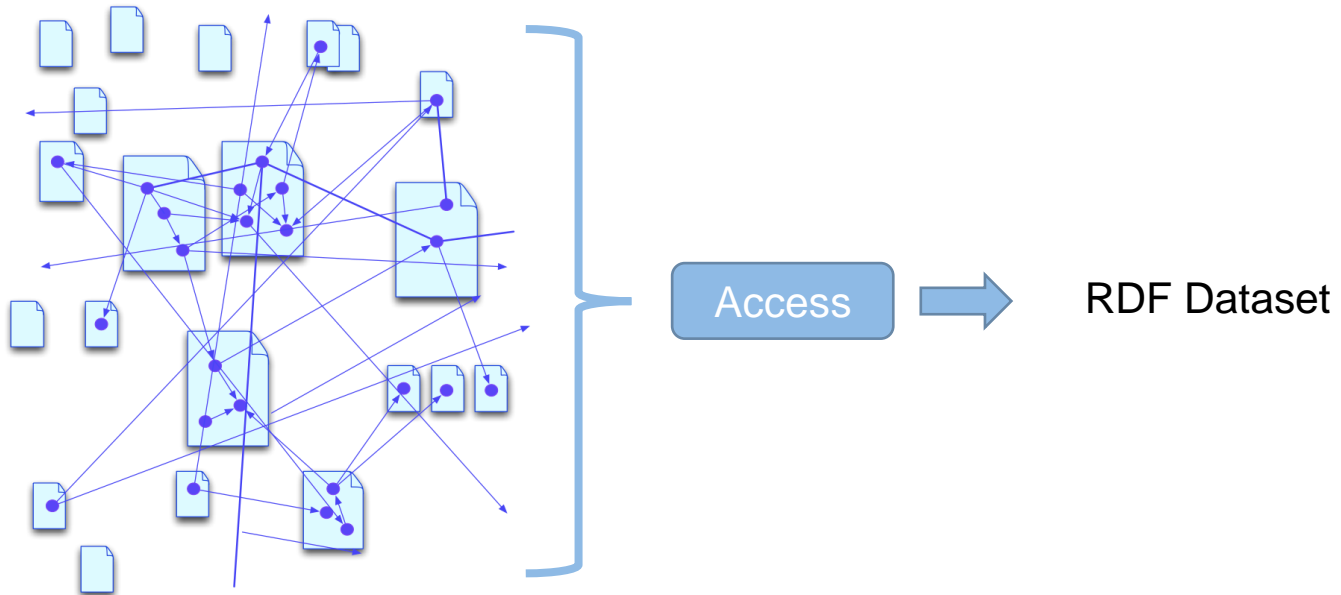Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY

Outlook – Chapter 2

- We will learn how to access Linked Data and construct a local RDF dataset out of Linked Data from the web
- We will consider how to merge RDF graphs, and how to crawl Linked Data

Access → RDF Dataset

[1] http://www.w3.org/DesignIssues/Abstractions.html

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN UNIVERSITY

Get the Idea?
(TOMORROW'S WEB WILL)

i
know
what
you
mean ...

PLUS:

Antibiotics'
Dim Future

Rorschach:
A Waste of Ink

The Oldest Stars

79

Semantic Web Exercises 1

General Information
The exercises are not graded and only serve as preparatory material for the presence exercise and exam. A solution will be presented in the exercise lectures.

E 1.1 Go to https://tools.ietf.org/html/ and select an RFC of your choice. Read it, and be able to summarize its content briefly during the next exercise.

E 1.2 What is the difference between IP-Addresses and Domain names? How does the DNS System work? What is an IRI? (Read RFC 3987)

E 1.3 Use a telnet client of your choice. Connect to the RWTH Main Website and request the main page from the Web Server. Does the RWTH Website set a Cookie? What kind of Cookie is it?

E 1.4
Read the RDF 1.1 Primer (https://www.w3.org/TR/rdf11-primer/).
In your own words:
• What is the relation between RDF, RDF vocabularies, XSD Schema and Turtle?
• Explain the terms resource, predicate, literal and IRI.
• Which requirements need to be fulfilled in order for a RDF graph to be semantically interpretable? Do you see potential issues?

Semantic Web Lecture
Prof. Stefan Decker
Lehrstuhl Informatik 5

Informatik 5 -
Informationssysteme
und Datenbanken

RWTH AACHEN
UNIVERSITY