



openHPI Web Technologies 2015 Week 1

Prof. Dr. Christoph Meinel
Hasso Plattner Institute
University of Potsdam, Germany

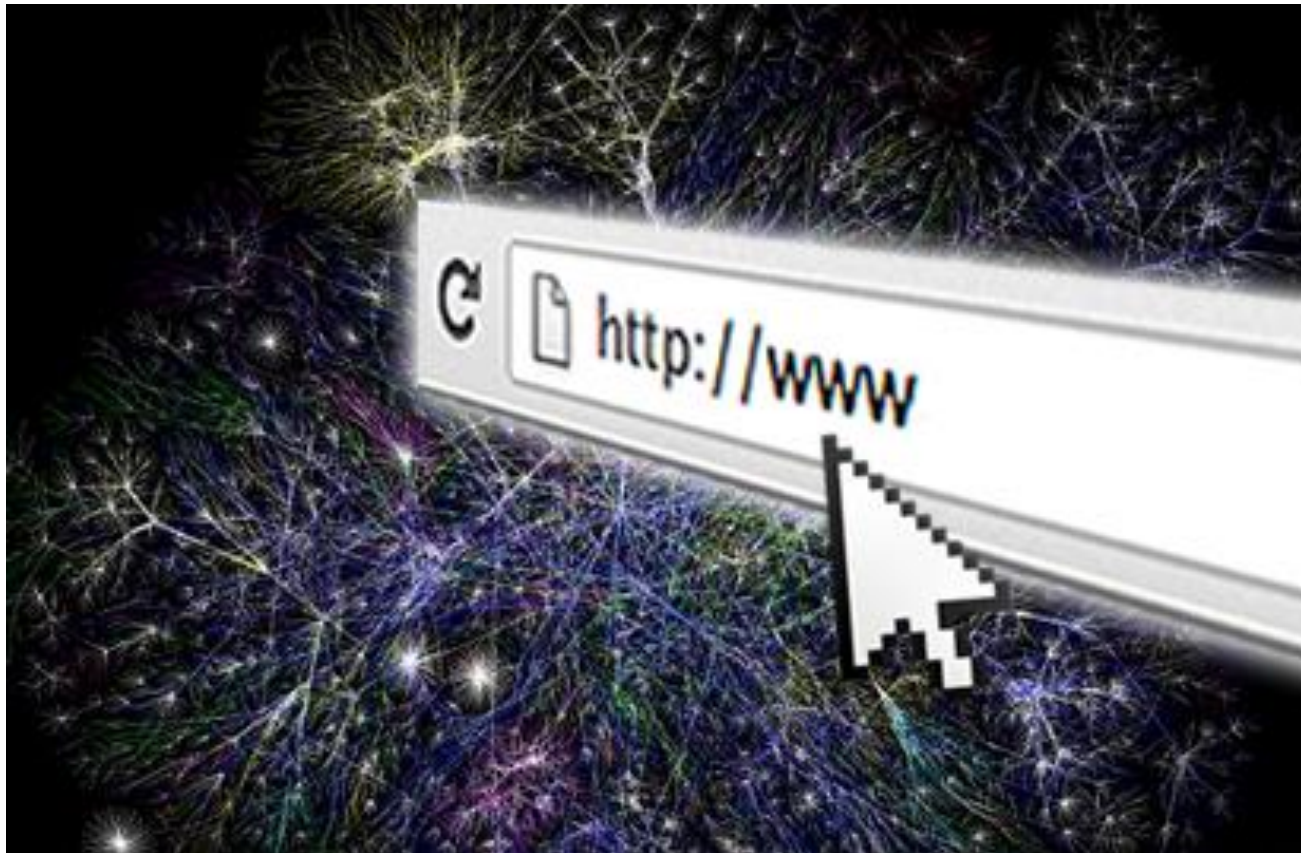


Overview of the openHPI Course Web Technologies

Prof. Dr. Christoph Meinel
Hasso Plattner Institute
University of Potsdam, Germany

WWW – World Wide Web, commonly called the Web ...

... everyone knows it. But how does it work?
Understanding this is extremely important in order to use it correctly.



Everybody who wants to understand how the World Wide Web works is welcome to the course

- No pre-requisites, course attendance is free
- Material is presented in form of learning videos and supplementary reading material
- We expect from our participants:
 - serious study of the provided learning material
 - self-control of individual learning progress via self-tests and homework
 - active participation in the discussion forums
- All those that complete homework exercises and pass the final exam will get a certificate

Week 1:

- Introduction to the Technology of the WWW

Week 2:

- URL and HTTP

Week 3:

- HTML, CSS and XML

Week 4:

- Client-side Web Programming

Week 5:

- Server-side Web Programming

Week 6:

- Social, Semantic and Service Web

Key Question:

*The Web has become an indispensable part of our daily life.
But what is its structure, and how does it work?*

Learning units:

- WWW – Introduction
- Hypertext, HTML and XML
- Browser and Server
- Web Applications and Web Programming
- Short History of the WWW
- WWW – Important Players
- Standardization in the Web

Key Question:

How can information and services on the Web be identified, and how can they be obtained via the Internet?

Learning units:

- Introduction URI
- URL and URN
- Introduction HTTP
- History of HTTP
- Content Negotiation
- Authentication and HTTPS
- Persistent Connections, Caching and Cookies
- HTTP Server Configuration

Key Question:

How can web documents be created? How can they be designed in such a way that not only we people can use them, but that they can also be processed automatically?

Learning units:

- HTML Documents
- History of HTML
- HTML Principles
- Dynamic HTML
- HTML 5 – State of the Art
- CSS, CSS Syntax, CSS Preprocessors
- XML

Key Question:

How can offers “come to life” on the Web? How can web applications be developed with HTML, JavaScript and libraries?

Learning units:

- Introduction to Web Programming
- Client-Side Web Programming
- DOM and DHTML
- JavaScript
- JavaScript Frameworks
- Selected JavaScript Libraries
- Example Application
- CoffeeScript and Dart

Key Question:

What kind of software is running on the server-side, and how can it be implemented with the help of web application frameworks, MVC, etc.?

■ Learning units:

- Introduction to Server-side Web Programming
- Session Management
- MVC
- Persistence and Database Layer
- Asynchronous Client-/Server Communication (AJAX, JSON)
- Web-based APIs and Webservices (REST)
- Graphics- and Video Formats, Character Encoding
- Selected Web Application Frameworks, Example Application

Key Question:

How can resources - computers, storage, software – be used efficiently on the Web?

Learning units:

- Introduction: Web 3.0
- Social Web and Social Media
- Facebook, Google+, Twitter / Renren, Qzone, Weibo & Co.
- Search Engines
- Semantic Web
- Machine-to-Machine Communication
- SOAP-based Webservices

Our Teaching Team Wishes You
Many New Learning Perspectives and a Lot of Fun !

The Course Teaching Team:



Prof. Dr. Christoph Meinel



Christian Willems



Matthias Bauer



Tom Staubitz



Dr. Haojin Yang



Martin Malchow



Xiaoyin Che



Cheng Wang



openHPI WWW – A First Approach

Prof. Dr. Christoph Meinel
Hasso Plattner Institute
University of Potsdam, Germany

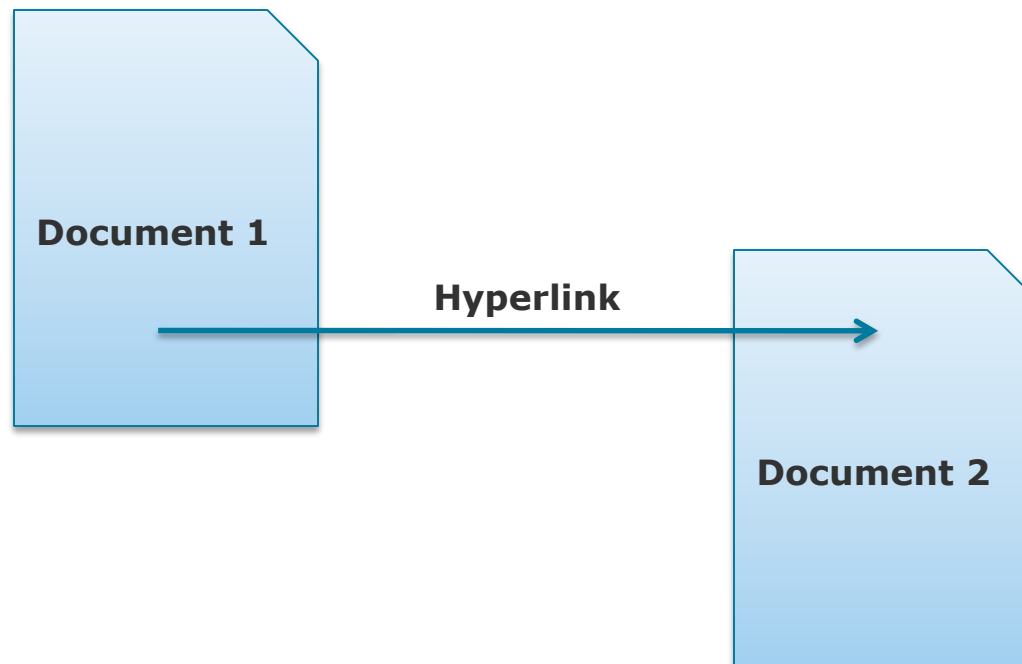
WWW – A First Approach (1/6)

- **World Wide Web, Web, WWW, ...** is a huge, globally distributed collection of information and data that can be accessed via the Internet with the help of **HTTP – Hypertext Transfer Protocol**
- The WWW has made the **Internet** the most important communication medium of our days
- Thanks to the intuitively usable, graphical interface – the **WWW browser** – the WWW provides simple access to its information offer to everybody

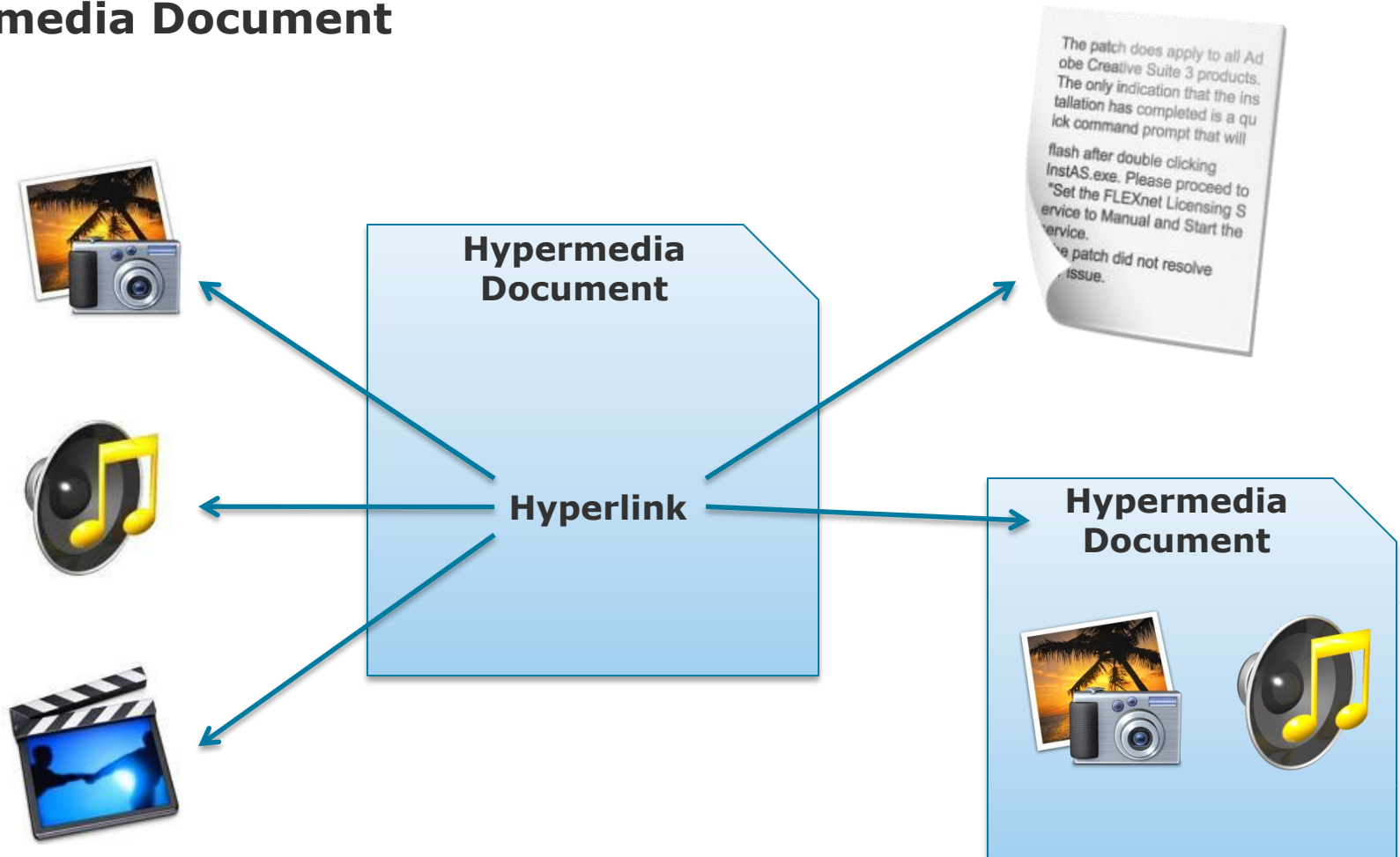
WWW – A First Approach (2/6)

- The WWW was introduced at CERN, the European Organization for Nuclear Research by Robert Cailliau and Tim Berners-Lee, at the beginning of the 1990s
- The information contained in the WWW takes the form of **hypermedia** documents (**hypertext documents**)
- Hypermedia documents are interconnected by so-called **hyperlinks** and in this way form an information network
- Users can navigate along these links quite simply through the gigantic network of web documents

Principle of the Hyperlink



Hypermedia Document



WWW – A First Approach (5/6)

- WWW documents are distributed worldwide, stored at so-called **WWW servers**
- In order to be found, they have to be uniquely identified globally
- Access to the WWW document proceeds according to the **client/server principle**:
 - the WWW client “browser” requests a specific document that is delivered by a WWW server
 - transmission through the Internet proceeds with the Hypertext-Transfer-Protocols (HTTP)

WWW – A First Approach (6/6)

- Modern browsers offer not only access to WWW documents but also to other Internet services, e.g.
 - FTP, email (as → Web application), streaming media (RTSP), ...
- The mechanisms and protocols necessary in order to carry out this service remain almost completely hidden from the user
- Thanks to the intuitive and simple to use browser-user interface, the WWW experienced tremendous growth and has now become the universal communication medium of our time



openHPI Hypertext, HTML and XML

Prof. Dr. Christoph Meinel
Hasso Plattner Institute
University of Potsdam, Germany

Hypertext and Hypermedia (1/2)

- WWW is a **hypermedia system** or **hypertext system**, WWW documents are interconnected to each other via hyperlinks
- The traditional linear structure of documents is superseded by hypermedia documents – via the linking mechanism they form a huge information network
- Users can directly access one WWW document from another, even when that document is stored on another server
- WWW is a **distributed** hypermedia system - WWW documents reside on servers distributed around the world. This makes it, for example, very difficult to guarantee consistency of the hyperlinks

Hypertext and Hypermedia (2/2)

WWW documents are composed in a special descriptive language:
the **Hypertext Markup Language - HTML**

With HTML, authors can

- describe the **structure** of a WWW document:
division into headings, paragraphs, tables,...
- include hyperlinks
- embed multimedia components
- influence (to a certain degree) the graphical presentation

HTML Documents (1/4)

- An HTML document is called a “**page**”
- The starting page or output page for navigation through the information offer of a provider is called the “**homepage**”
- Browsers interpret the requested HTML documents, prepare them and display them based on HTML identifiers in the document, the so-called “**markups**” or shortly “**tags**”
- **Markup language** should only be used for the description of the document structure and not for description of the document presentation

HTML Documents (2/4)

Every HTML document consists of two parts:

- document “**header**”
 - contains information about the document
- document “**body**”
 - contains the actual contents of the document

HTML Documents (3/4)

Basic idea of the markup language was the **separation** of structure and design

- **HTML** describes the document structure
- **Cascading Stylesheets – CSS** – are responsible for the description of the presentation (layout) of this structure
 - Main area of application for CSS:
 - exact definition of the layout of an HTML document
 - adoption to various output media
 - central layout management

HTML Documents (4/4)

- With the help of special tags, reference “links” can be stored in an HTML document as passive “pointers” to another document
- So-called **URIs – Uniform Resource Identifiers** - serve in the unique global identification of documents
- URIs can be encoded in strings and are made up (in the case of URLs) of:
 - name of the access protocol,
 - computer name and
 - document name

- The main criticism of the markup language HTML has been its lack of flexibility
- Through the introduction of **XML – Extensible Markup Language** – as meta-markup language, it is possible to describe each type of document with its own syntax
- XML is the starting point for a variety of application-specific markup languages, that are tailored exactly to the characteristics of specialized output devices or application areas, e.g.
 - MathML – Mathematical Markup Language
 - SVG – Scalable Vector Graphics
 - ODF – Open Document Format for Office Applications
 - ...



openHPI Browser and Server

Prof. Dr. Christoph Meinel
Hasso Plattner Institute
University of Potsdam, Germany

WWW Browser

- WWW users can access the huge supply of information and data with the help of **browsers**
- Browsers today offer a **Graphical User Interface, GUI**
- With simple “point and click” actions, users can carry out a variety of functions and retrieve WWW documents
- The browser interprets the WWW documents and presents them in processed form: texts, graphics, images, videos, audio, ...

Widely used browsers:

- Chrome, Internet Explorer, Firefox, Safari, Opera, ...

Browsers are

- usually free and available for all platforms – from smartphone to mainframe
- hardly differ in user guidance

Browsers (still) often behave slightly different in their support of certain WWW standards. This stems from the 1990s when browser manufacturers tried to improve their competitive edge by offering their own proprietary functionalities and extensions of the WWW

→ “**browser war**”

Client / Server Paradigm

Access to WWW documents is organized according to the **client/server paradigm**:

- User requests a WWW document over the WWW client browser (by mouse click or by entering a URL)
- Browser contacts the WWW server specified by the URL and requests the desired document
- Server accesses its local file system and sends the file specified in the URL to the requesting host
- Browser receives the document
- Browser interprets the document and displays it

HTTP – Hypertext Transfer Protocol

- Interaction between browser and server is carried out by means of the **Hypertext Transfer Protocol - HTTP**
 - HTTP is a very **simple** and **stateless** (no “memory” of earlier data exchange) – therefore fast – protocol, interaction proceeds in the form of a simple question / answer procedure
- Efficiency of HTTP communication can be increased significantly with the help of **intermediate systems** such as
 - proxy servers
 - gateways
 - cache storages

HTTP – Hypertext Transfer Protocol

HTTP is subject to a developmental process

- First HTTP version (HTTP/0.9) emerged in 1989/90 at CERN
- Current version is **HTTP/1.1** (since 1999)
- At this time **HTTP/2.0** is in the process of being standardized
 - based on the research project SPDY (“Speedy”) which was initiated by Google
 - it will probably be finished this year

Securing HTTP Communication

There are two ways to make HTTP communication secure:

- **TLS – Transport Layer Security**, and its predecessor **SSL – Secure Sockets Layer** – provide secure communication infrastructure for HTTP communication
- **S-HTTP** (Secure HTTP) offers an independent communication protocol based on HTTP (but has not been able to establish itself and so it is barely used)

WWW Server

Task and architecture of a **WWW server** is quite simple:

- server waits for the opening of a connection through a browser
- server answers browser's request
- after sending the answer, the server ends the connection and waits for new requests
- WWW servers today fulfill many tasks beyond this:
 - session management (Cookies)
 - script interpreter for dynamic web applications
 - ...

Web Browser Architecture

Tasks and architecture of a **browser** are considerably more complex:

- after user input, opening of a connection to the selected server
- reading the information resource received by the server
- processing and presenting the information resource received by the server
- response to the user reaction on operating the graphical user interface (GUI)

Web Browser Architecture

To solve all tasks, the browser consists conceptually of multiple clients and interpreters as well as a controller

Various browser components (1/3):

- The **controller** is the central instance of the browser architecture:
It manages and coordinates the actions to be performed, e.g.
 - interprets mouse clicks and keystrokes
 - activates other components of the browser for performing requested operations

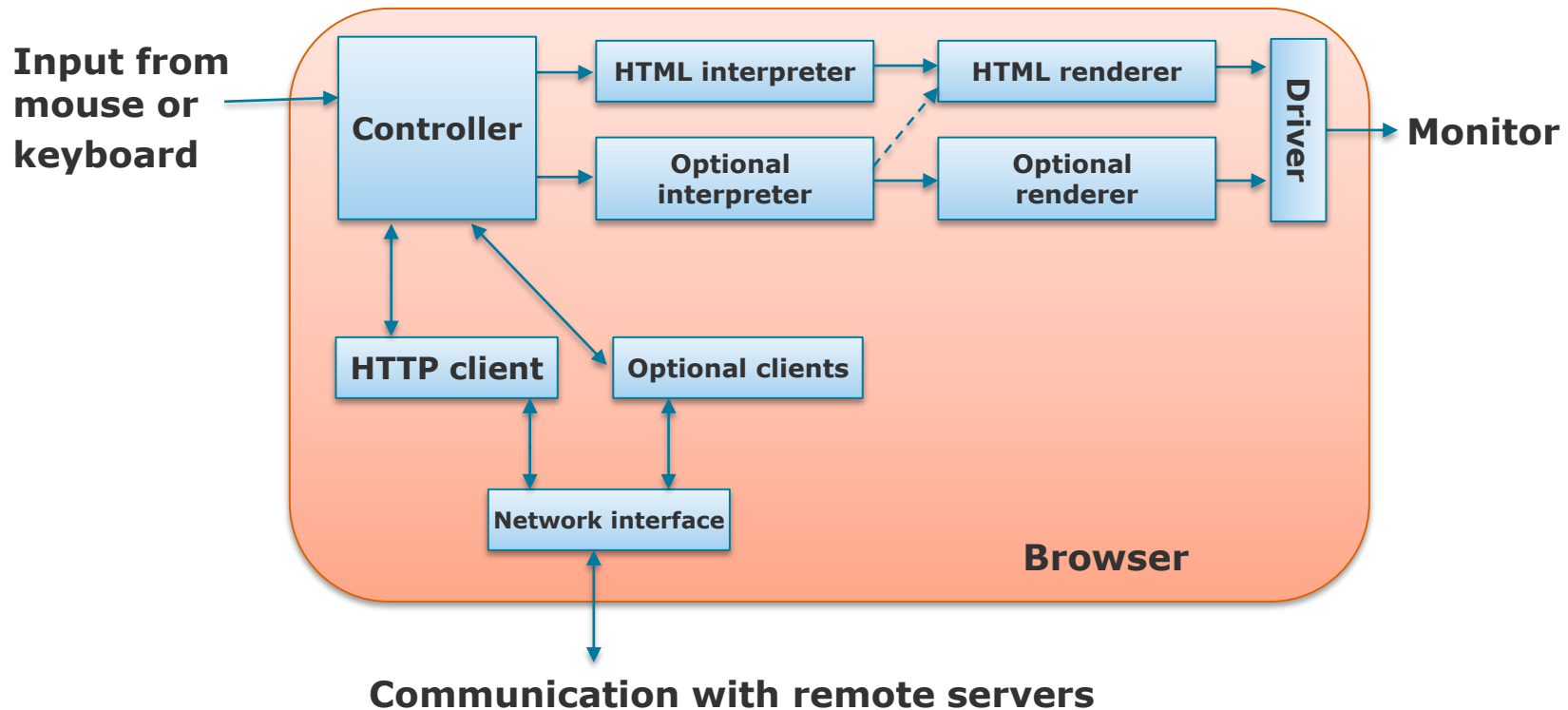
Various browser components (2/3):

- The **HTML interpreter** serves to display the user-requested HTML documents correctly
 - interpretation of the HTML tags in the HTML document
 - processing the HTML documents for presentation
 - on-screen display
 - correct interpretation of the user-selected active elements of the HTML document (management of the relationship between display and anchored element)

Various browser components (3/3):

- The **Cache** (a kind of memory) serves to increase the performance and throughput of a browser with intermediate storage of the already visited HTML documents
 - Browser puts a copy of the read HTML pages in the cache
 - If the page is requested again, the saved copy is displayed, when ensured that the copy is still valid ...

Web Browser Architecture



Browser enhancement with plug-ins

- Plug-ins allow the browser to process and display data that does not correspond to the typical formats. e.g.
 - HTML, CSS, pictures, etc.
- Prominent examples:
 - Java (Applets), Adobe Flash, Microsoft Silverlight, PDF, various audio and video formats
- Plug-ins vs. extensions
 - Browser plug-ins only allow the display of plug-in specific contents in web pages, i.e. enhance interpreter and rendering components
 - Extensions (also: add-ons) modify or extend the functionality of the browser itself, e.g.: ad blocker, download helper, ...



openHPI Web Applications

Prof. Dr. Christoph Meinel
Hasso Plattner Institute
University of Potsdam, Germany

Web Applications

Today, the Web is no longer just an information archive, but has become a marketplace for a wide range of (Web) applications, e.g. shops, mail clients, dating sites, ...

Branch software or office packages (e.g. Google Docs) not only deliver static documents but also

- allow interaction with the user
- are able to process data from users
- generate documents dynamically dependent on
 - user input (in forms)
 - inquiries with parameters (e.g. search queries)
 - events (time, weather, etc.)
- Distinguish between: **client-side** and **server-side** web applications
➔ different types of web programming

Client-Side Programming

- Executable / interpretable program code is transferred from the WWW server within the requested HTML document to the WWW browser (client) and processed at the client computer
- WWW client needs interpreter (if required as a plugin) which, together with the client, can interpret and execute the transferred program code
 - → Java applets
 - → JavaScript program (or script)
 - → Microsoft Silverlight and Adobe Flash

Server-Side Programming

- Application program is activated by the WWW browser (client) via the WWW server at the server computer and executed there
- Processed results are given to the client as a dynamically generated HTML document
- Standardized interface between WWW server and application programs at the server side: **CGI interface**
- Application programs at the server side can be written in desired programming language, like
 - Java (Servlets, Server Pages, Beans)
 - script languages (ASP, PHP, Perl)
 - today: Web frameworks (Ruby on Rails, Spring, Django, ...)



openHPI
Short History of the WWW

Prof. Dr. Christoph Meinel
Hasso Plattner Institute
University of Potsdam, Germany

Short History of the WWW

The history of the WWW and the Internet are inseparable

- 1969: Start of the Internet:
 - ARPANET started with 4 computers at universities in Los Angeles, Santa Barbara, Stanford and Utah
- 1971: 23 hosts are linked in the ARPANET via 15 nodes
- 1971: First email sent (killer application ARPANET)
- 1972: Specification of the File Transfer Protocol (FTP)
- 1973: Specification of the protocol for sending email
- 1973: First international nodes in England and Norway

Short History of the WWW

- 1983: Already over 500 connected hosts
 - Changeover of communication software to TCP/IP, developed under the direction of Vinton Cerf and Robert E. Kahn (birth of the Internet)
 - Separation of the ARPANET into civilian and military areas
- 1986: NSFNET program supports Internet connection of all universities in the US
- 1988: First Internet worm, with 10% of the 60,000 hosts affected
- 1989: 150,000 hosts are connected with the Internet, ARPANET is shut down

Short History of the WWW

- 1990: Development of the first WWW browser and WWW server by Robert Cailliau and Tim Berners-Lee (birth of the WWW)
- 1991: First version (0.9) of the Hyper Text Transfer Protocol (HTTP)
- 1993: First browser - NCSA Mosaic – with graphical user interface
- 1994 – Founding of the W3C, founding of Netscape
- 1995: MS provides Windows 95 with the browser Internet Explorer
- 1995: Finance and business discover WWW, creation of many “dot-coms”
- 1999 1st generation mobile Internet: WAP, i-Mode in Japan

Short History of the WWW

- 2001: The “dot-com bubble” bursts
(boom of young Internet companies collapses)
- 2001: Wikipedia founded
- 2002: The Web is everywhere: successful Web offers crystallize from the remains of the “dot-com bubble”, among them Google, Amazon, eBay, various airlines
- 2003: Advent of the social networks
 - Myspace (2003): initially >200,000 users daily, insignificant today
 - Facebook (2004): launch in February, 1 million users in December 2004, 1.39 billion active users in February 2015

Short History of the WWW

- 2004 – UMTS commercially available
- 2005 – “Web 3.0”: Research in the areas of Semantic Web, Social Web and Service Web
- 2007 – The first iPhones appear, touchscreen revolution
Smartphone market → breakthrough of mobile Web
- 2008 – Android officially available for the first time
- 2009 – “Cloud Computing” becomes an important trend in the coming years → Processing power and memory in the Web and access to these resources with mobile devices introduces paradigm shift
- 2010 – First iPad hails renaissance of the tablet PC
- Today– Web is ubiquitous: mobile Internet widespread, smart homes, smart grid, ..., important concepts for the future



openHPI Players on the Web

Prof. Dr. Christoph Meinel
Hasso Plattner Institute
University of Potsdam, Germany

Who's Who in the WWW

Internet and WWW have no central control authority

- Internationally composed, non-profit organizations interact with the entire Internet community to regulate the technical organization and development of the Internet in form of a public standardization process

IAB – Internet Architecture Board – determines the direction and areas of the Internet for further development, and organizes them

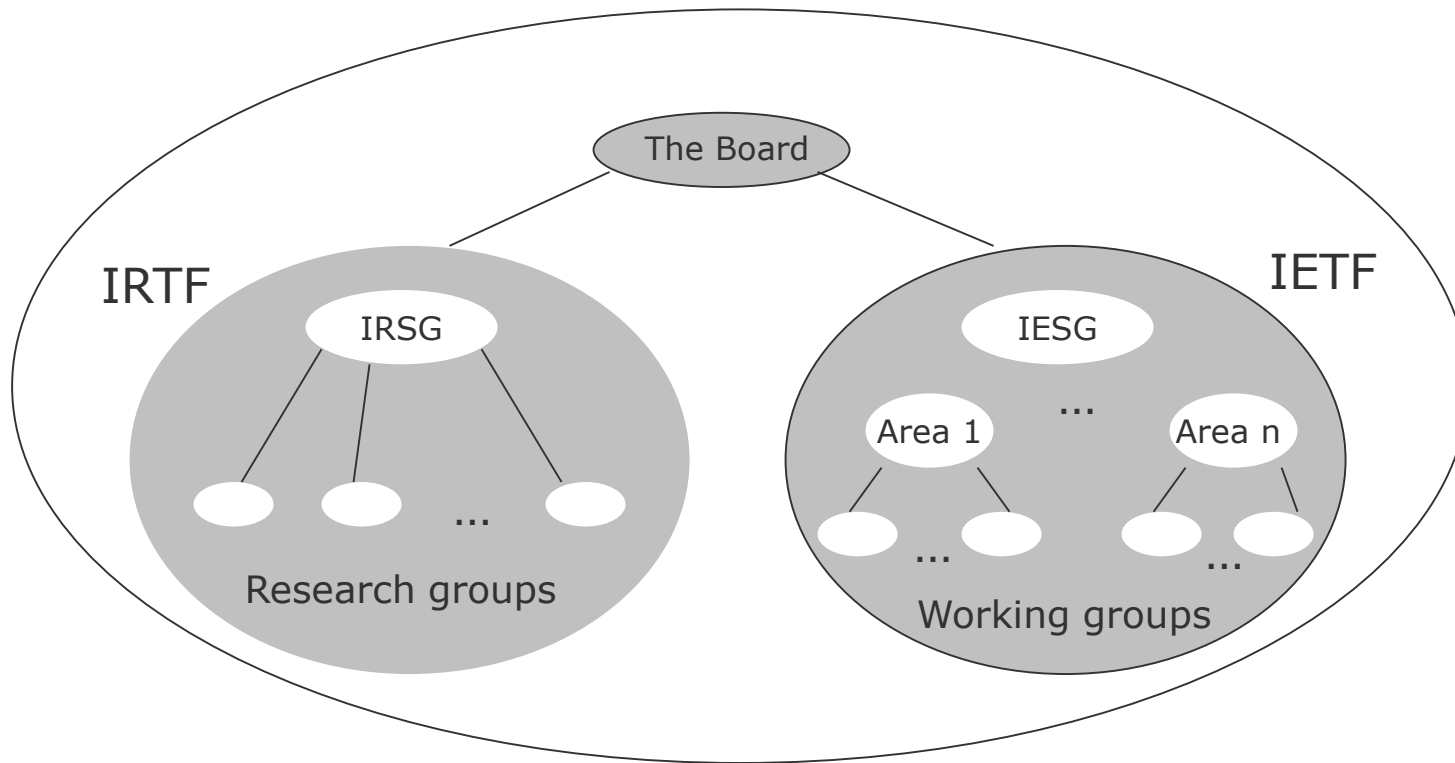
- IAB organizes itself via so-called **Internet Task Forces**
 - Here, special questions are dealt with and status reports are submitted to IAB

Who's Who in the WWW

Two support groups have been formed for the **IAB**:

- **IETF** – Internet Engineering Task Force
 - is led technically by **IESG** – Internet Engineering Steering Group
 - is concerned with the standardization of communication protocols in the Internet
- **IRTF** – Internet Research Task Force
 - is concerned with research questions on Internet technologies

Structure of the IAB - Internet Architecture Board



Who's Who in the WWW

- **IANA** – Internet Assigned Number Authority
 - until 1998 responsible for issuing addresses on the Internet,
 - In 1998 the responsibility shifted to ICANN – Internet Corporation for Assigned Names and Numbers
- In 1992 the **ISOC** - Internet Society was organized as an international social superstructure for this organization
 - Most important task: Setting new standards

W3C – World Wide Web Consortium

- Founded in 1994 at MIT by Tim Berners-Lee
- Originated from the WWW working group of the IETF as this group could not publish any standards of its own
- W3C develops and maintains standards that apply to the WWW, e.g. HTML, CSS, XML, SVG, OWL, etc.
- W3C is not an officially recognized organization, and, therefore, not able to set official standards (e.g. ISO norms)
 - ➔ W3C publications are, therefore, called *recommendations*, although they are actually often de facto standards
- W3C finances itself through membership fees;
 - since 2009 it has been financially supported by ISOC through donations



openHPI Internet Standards

Prof. Dr. Christoph Meinel
Hasso Plattner Institute
University of Potsdam, Germany

Internet Standards

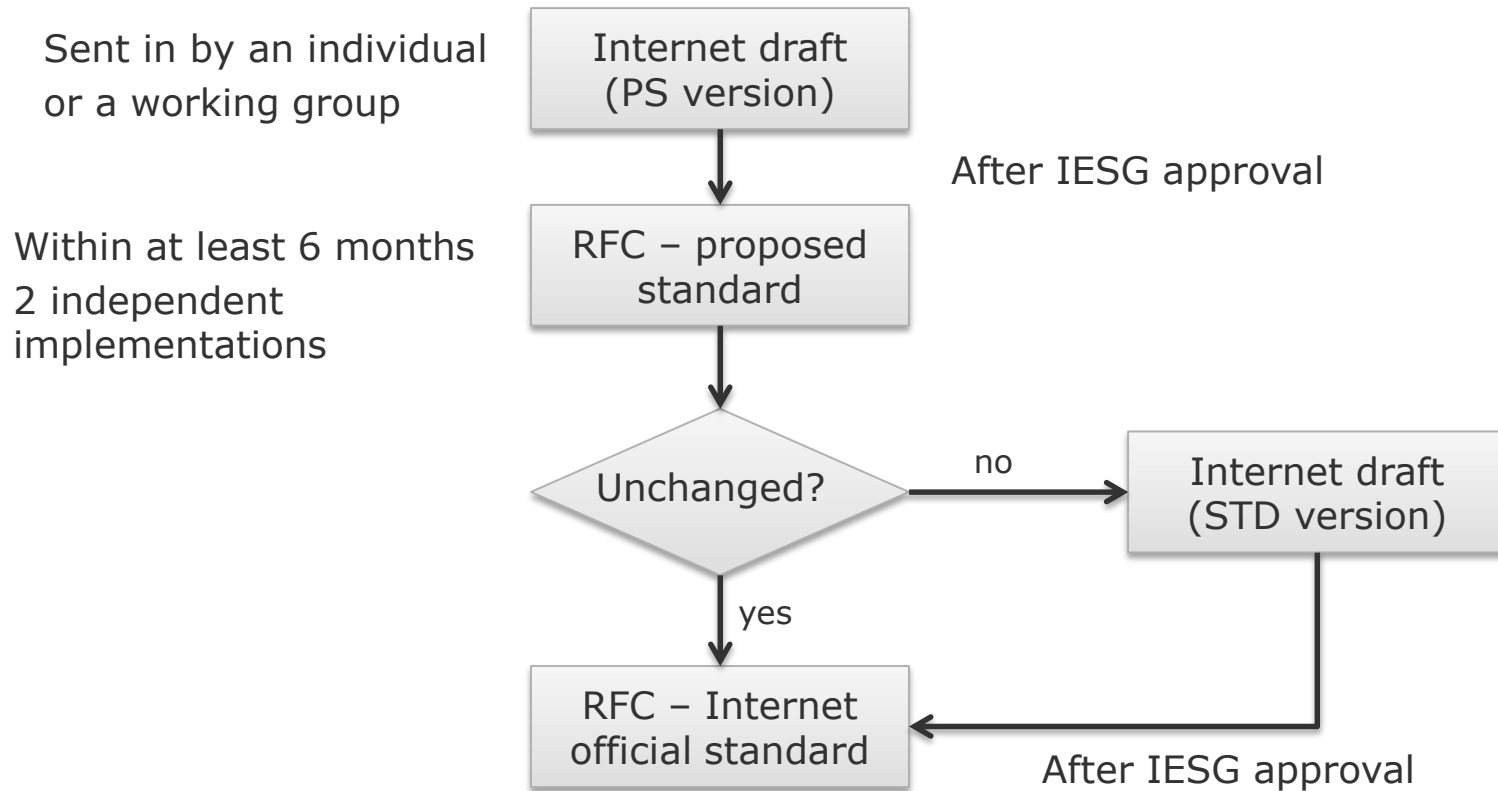
- **Internet standard** is a specification that is supported and used by the Internet community
- Starting point for Internet standards are **RFC – Request for Comments**
- RFCs contain proposals for establishing new Internet standards or revising existing ones
- In order to include the Internet community, RFCs are published online free of charge by the IESG or IAB

Standardization process (maturity levels):

- **Internet draft** – submitted by individual or working group
- **RFC proposed standard** – drafts approved by IESG IETF
 - Pre-requisite: For proposed standards, at least two independent and interoperable implementations within 6 months must exist
- **RFC Internet Standard** – after sufficient experience with implementations

IAB establishes a “status” for every standard (required, recommended, elective, limited use, not recommended) which characterizes its meaning

Standardization process on the Internet



W3C Recommendations

- Development of W3C recommendations similar to IAB standards, but somewhat less restrictive
- Stages of development:
 - Working draft: Openly accessible and open to comment
 - Candidate recommendation: Only minor changes, development of implementations
 - Proposed recommendation: Supplemented by documentation and evaluation of implementations
 - After approval by W3C committee: Recommendation
- All W3C drafts and recommendations are available anytime online