



The O'Reilly Peer-to-Peer & Web Services Conference

Will Peer to Peer Unravel the Web?

Noah Mendelsohn

IBM & Lotus Development Corp.



The Universal Web

“By Universal I mean that the web is declared to be able to contain in principle every bit of information accessible by networks. It was designed to be able to include existing information systems such as FTP, and to be able simply in the future to be extendable to include any new information system.”



Tim Berners-Lee
December, 1996

The problem...



What I want...

Bearshare/
Gnutella

HTTP
download

Napster



P2P and the Web

- **Some P2P systems integrate w/the Web**
 - Dave Winer's "two-way-web"
 - Freenet (URI's)
 - Magi
 - etc., etc.
- **Some important ones don't!**
 - Napster
 - Gnutella/Bearshare
 - Etc., etc.



Consequences

- **You can't link**
- **You can't integrate information**
- **Client tools inappropriate coupled to storage mechanisms**
- **You can't move resources between web & P2P -- or across P2P “worlds”**
- **Naming and query gets reinvented**
 - Google crawls don't work
 - Gnutella can't find web resources



The solution...

...integrate P2P with the Web!



The web is not architecturally dependent on...

- **HTTP**
- **DNS**
- **IP V4***
- **Location-based naming**



* Actually, literal IP V6 addresses are not allowed in URI's

Peer-to-Web Roadmap

- **Invent URI schemes to name information stored in P2P systems**
- **Decouple browsers, music players, presentation software, application servers, etc. from storage mechanisms**
- **Deploy software in browsers, music players, app servers, P2P systems, etc. that can resolve and retrieve from a broad range of URI schemes.**



Naming



URI: Names for Web Resources

“A Uniform Resource Identifier (URI) is a compact string of characters for identifying an abstract or physical resource.”

*IETF RFC 2396
December, 1996*

“If something does not have a URI, you can't refer to it, and the power of the Web is the less for that.”

*Tim Berners-Lee
December, 1996*



What's a URI, do you mean URL?

- **Short answer:** you've got the right idea
- **Better answer:**
 - URI: the name for a resource on the web
 - URL: a URI that helps you *find* the resource



It's on my server

URL →

<http://noah.com/songs/yesterday.mp3>

<songs://Beatles/yesterday.mp3> ← URI



I have no clue where it is



URI Axioms*

- **Universality**
 - Any resource anywhere can be given a URI
 - Any resource of significance should be given a URI.
- **Global scope**
 - It doesn't matter to whom or where you specify that URI, it will have the same meaning.



*From: <http://www.w3.org/DesignIssues/Axioms.html>

URI Axioms (cont')

- **Sameness**
 - a URI will repeatably refer to "the same" thing
 - the significance of identity for a given URI is determined by the person who owns the URI, who first determined what it points to
- **Non-unique:**
 - URI space need not be the only universal space
- **Opacity**
 - The only thing you can use an identifier for is to refer to an object {...} you should not look at the contents of the URI string to gain other information.
- Also: **cool URI's don't change**



The “scheme”
tells how to
handle the rest
of the name

What's in a (URI) name?

`http://noah.com/somedoc.html`

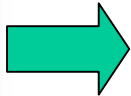
`mailto:Noah_Mendelsohn@lotus.com`

Browsers and web servers come with software modules to handle particular URI schemes (`http:`, `ftp:`, `mailto:`, etc.)



How about URI's for P2P?

`http://gnutella.com/beatles/yesterday.mp3`



`napster://songs/beatles/yesterday.mp3`

`songs://beatles/yesterday.mp3`



Variations on P2P URIs

napster://songs/beatles/yesterday.mp3

A resource in the Gnutella space implements a keyword query

...e/beatles+yesterday

~~Gnutella://keywordspace?beatles+yesterday~~



Challenge: managing URI schemes

However, we do not do this lightly. When we introduce a new space, it may have very different properties and we expect that the deployment of new software will be needed to allow access to it...If we use this extension point frivolously, ironically, it will cease to work.

Suppose very many schemes are introduced. The access scheme space itself becomes a namespace with all the problems which current namespaces such as DNS are trying to solve....

*Tim Berners-Lee
December, 1996*



The //localhost hack

`http://localhost:9876/yesterday.mp3`

- Trying to integrate with the web, but...
- Only works where software is installed at the port
- Inappropriately uses http as gateway protocol
- Uses port as pseudo-scheme



The //localhost hack

`http://localhost:9876/yesterday.mp3`

Violates URI Global Scope Axiom:

“It doesn't matter to whom or where you specify that URI, it will have the same meaning.”



Query vs. naming?

- **It's a duality for both the web and P2P**



Query vs. naming?

- It's a duality for both the web and P2P
- Web:



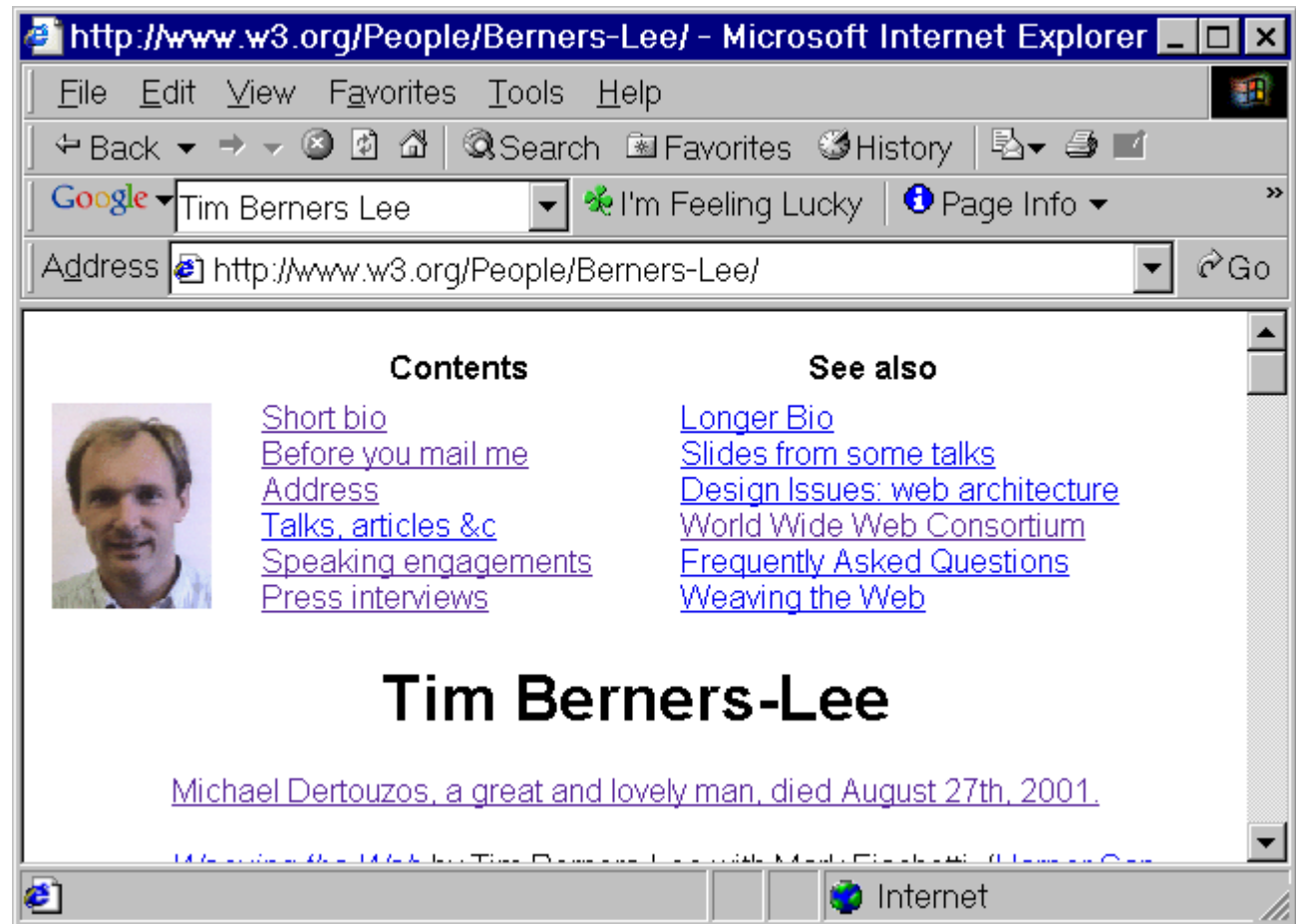
VS.



...both give you the same page



Query vs. naming



Query vs. naming?

- It's a duality for both the web and P2P
- Web:



VS.



...both give you the same page

- Many P2P systems are query-based
- Whichever style is chosen, the web and P2P can be integrated



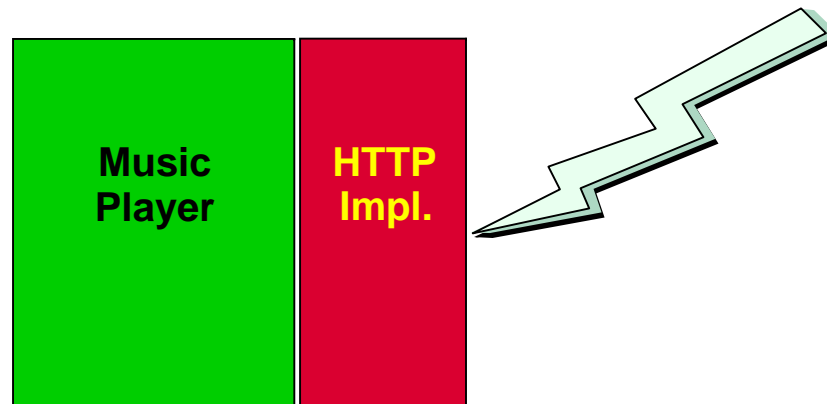
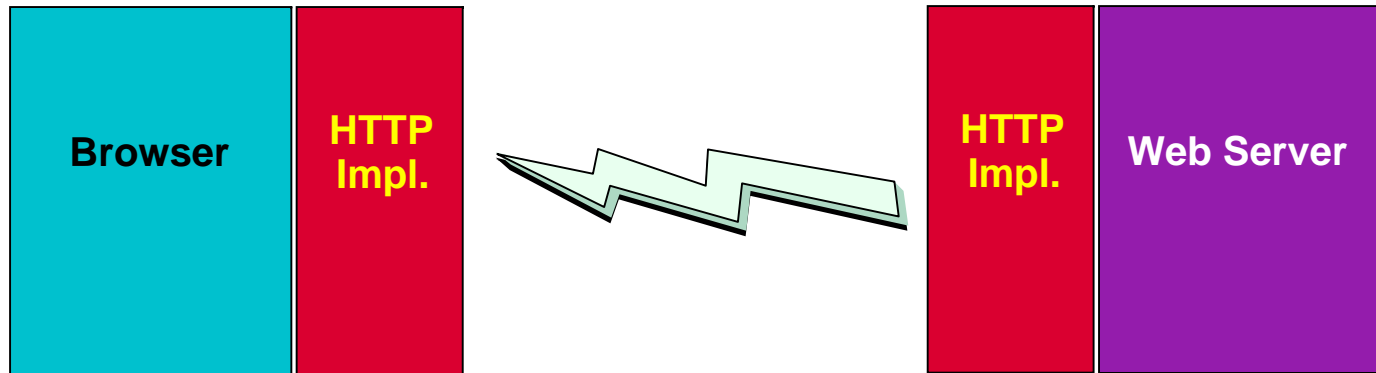
Restructuring Web and P2P Software



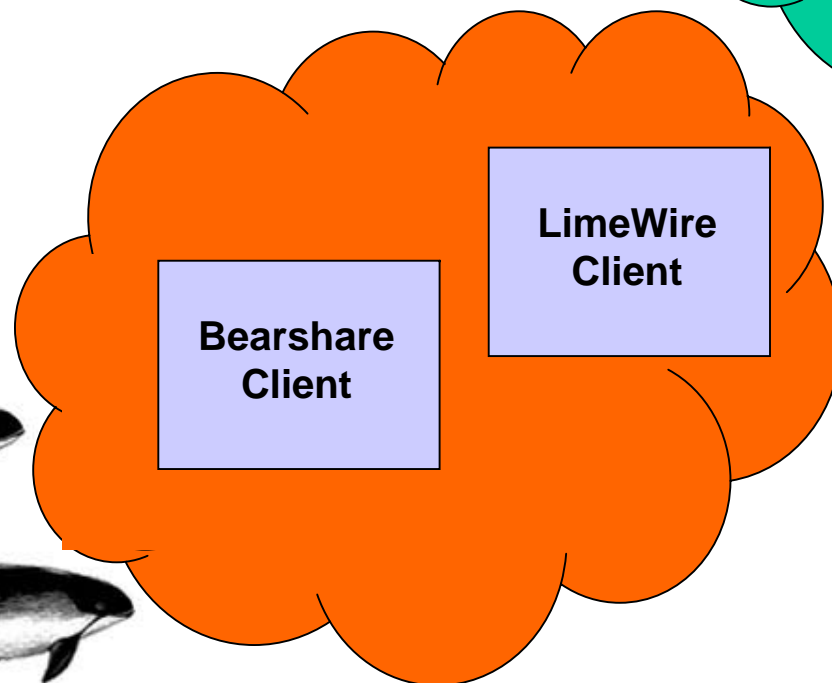
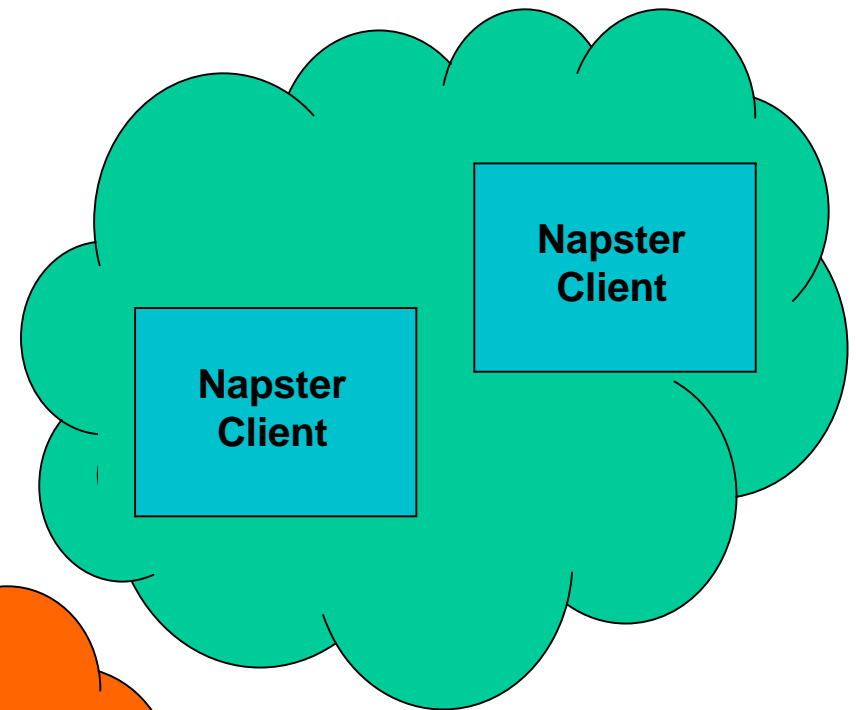
O'REILLY®

Will P2P Unravel the Web? – Noah Mendelsohn

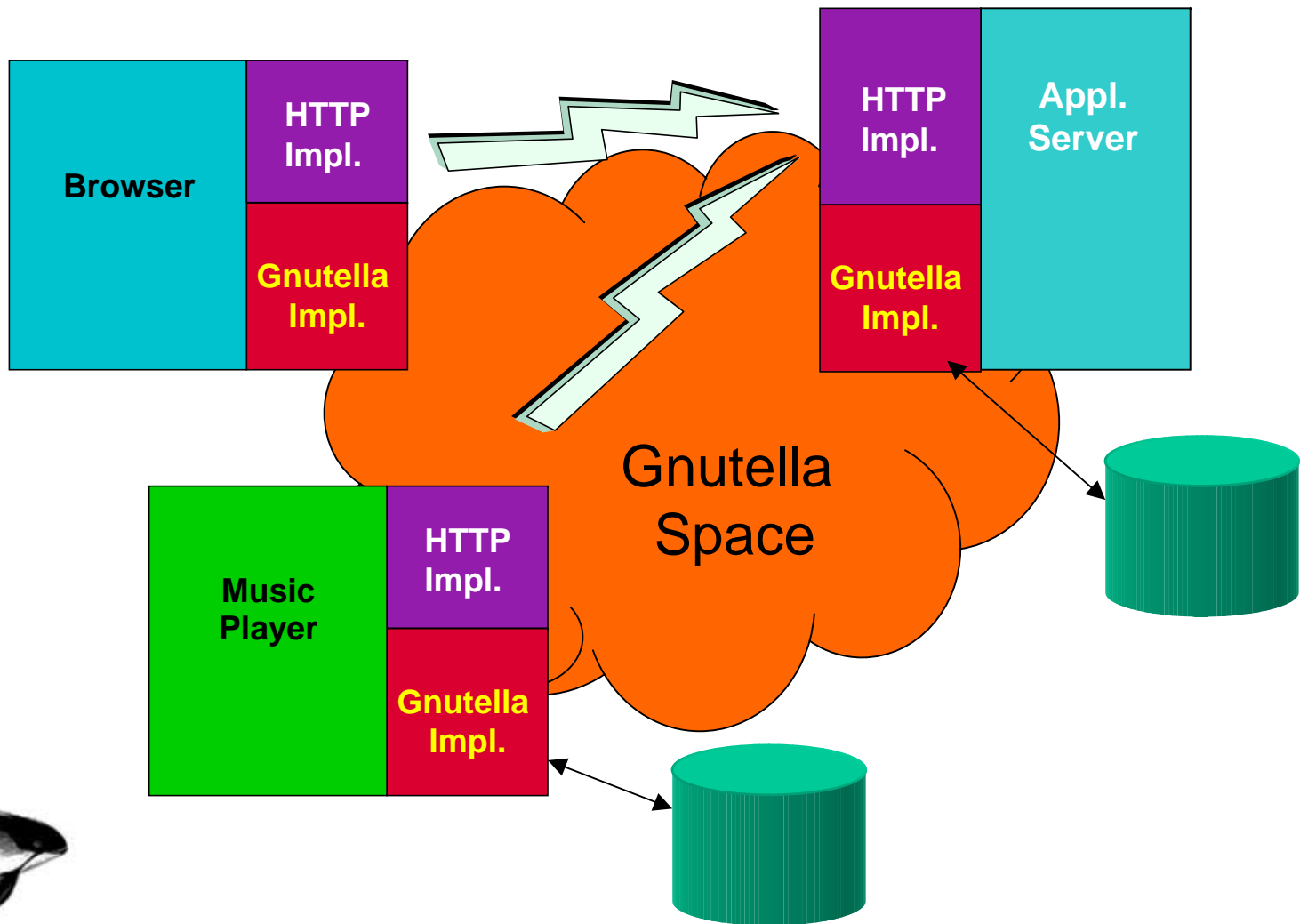
Web Today..



P2P Today



Possible Web Tomorrow



Conclusions



Summary

- **If P2P charts its own course, it could indeed “unravel the web”**
- **The web is big enough integrate P2P...there is only one web.**
- **Doing it:**
 - Getting naming right is key
 - Open up existing client, server & P2P software
 - Decouple “client” & app server software from particular storage & query models
- **It’s important! Let’s try.**



Some useful TimBL references

- **Web Architecture from 50,000 feet**
 - <http://www.w3.org/DesignIssues/Architecture.html>
- **Universal Resource Identifiers -- Axioms of Web Architecture**
 - <http://www.w3.org/DesignIssues/Axioms.html>
- **Guide to other web design documents**
 - <http://www.w3.org/DesignIssues/Overview.html>
- **Cool URI's don't change**
 - <http://www.w3.org/Provider/Style/URI.html>



Another reference

- Uniform Resource Identifiers (URI): Generic Syntax (IETF RFC 2396)
 - <http://www.ics.uci.edu/pub/ietf/uri/rfc2396.txt>

