

Measuring Open Source Popularity

Luke Welling

Hitwise

luke.welling@hitwise.com



First step is admitting your problem

- Hi, my name is Luke and I am a data addict ...



Measurable Data

- Sales figures?
- Downloads?
- Automatically interrogate
- Visits to web sites
- Traffic on Mailing Lists
- Books on Amazon?
- Search engine activity?
- Blogs/Webpages?
- Version Control System activity?
- Oscon Attendances/job listings/T-shirts?

- Some of this data is public, some is not

Sales Figures?

- For proprietary software the most obvious metric to track for comparisons or trends is sales
- The level of public detail provided in press releases and annual reports varies, but internally it is an obvious crucial measure
- You are likely to see more public detail in good times than bad
- Open source applications generally don't have many or any sales

Downloads?

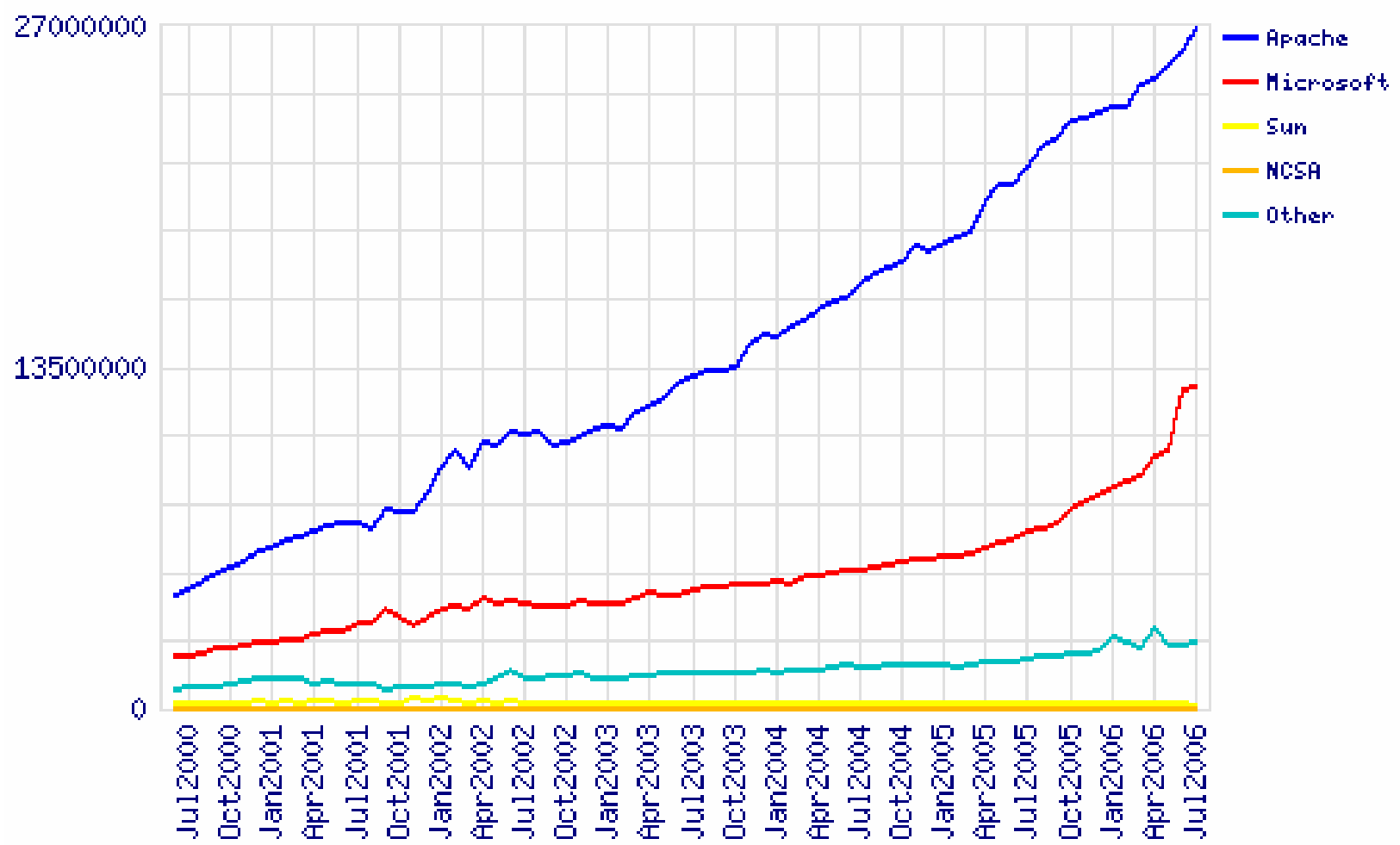
- Depending on their distribution models, an OS project might have a good idea of how often their application is downloaded
- Only some choose to ~~brag about them~~ make them public
- MySQL claim over 50,000 downloads a day*
- But again, you are likely to see more public detail in good times than bad
- Popular applications are only obtained directly a fraction of the time though

* Source: <http://www.mysql.com/why-mysql/marketshare/>

Automatically Interrogate

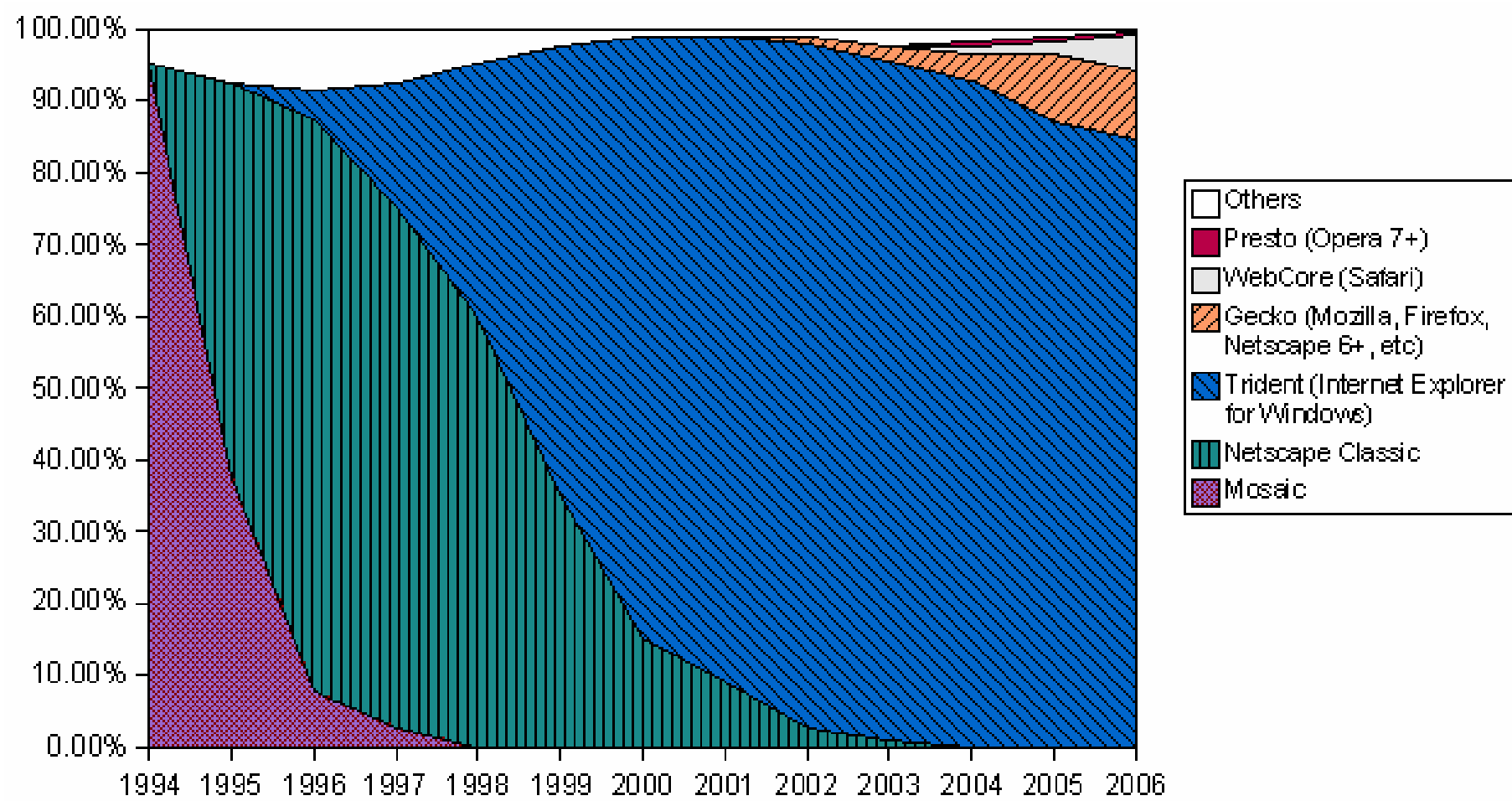
- Some types of software are routinely used in public
- It is practical to examine the signatures of web browsers and web servers and make conclusions from the sample you can see

Netcraft's Web Server Survey - Active



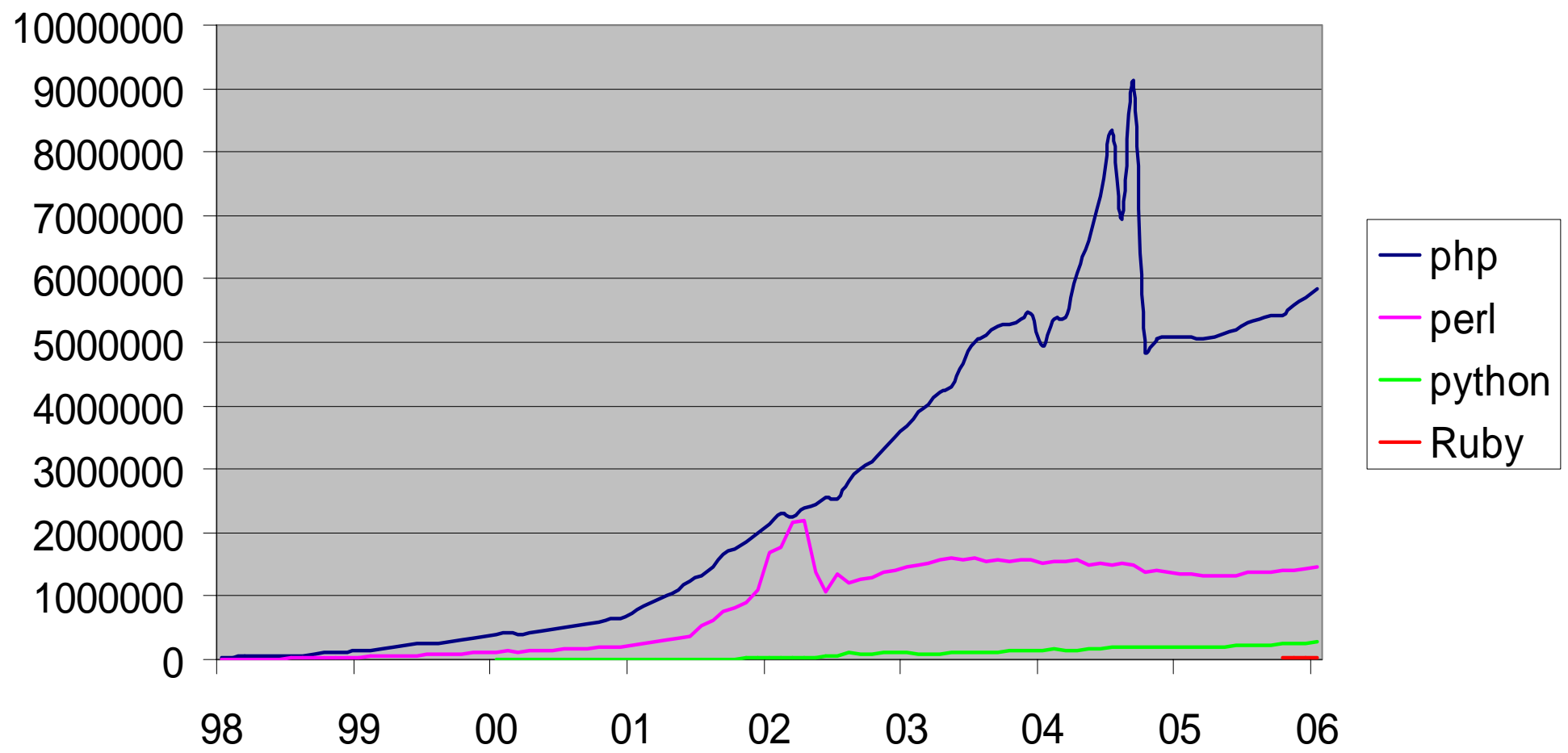
Source: <http://www.netcraft.com/>

Web Browser Usage Share



Source: http://en.wikipedia.org/wiki/Usage_share_of_web_browsers

E-Soft's Apache Module Survey



Source: <http://www.securityspace.com>

Web Site Visits

- A visit to an open source website could be for many things
 - Download
 - Documentation
 - Participation
- Directly comparing two sites' traffic may not always be meaningful.
- A site with a very good manual with user comments is going to get more traffic than a site with less documentation or less interaction
- Trends over time should be more telling

Web Site Visits

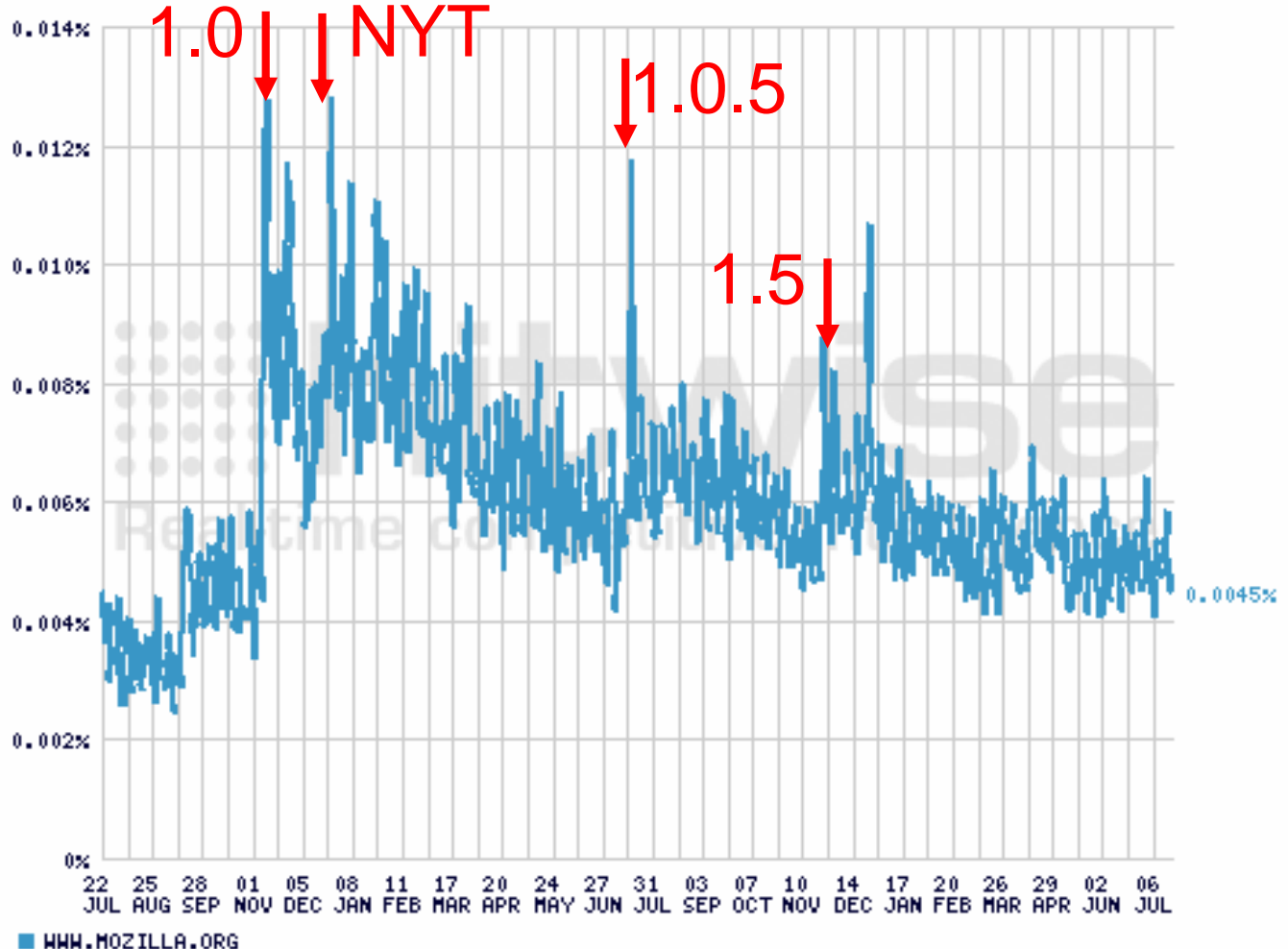


CHART OF THE DAILY ALL SITES MARKET SHARE IN 'ALL CATEGORIES', BASED ON VISITS.
TIME PERIODS REPRESENTED WITH BROKEN LINES INDICATE INSUFFICIENT DATA.
GENERATED ON: 07/22/2006. COPYRIGHT 2006 (C) 'HITWISE PTY LTD'.

You can see detail ...

- November 9, 2004 - Official Version 1.0 release
- December 16, 2004 – New York Times Advert
- July 12, 2005 – 1.0.5 Vulnerability patch
- November 29, 2005 - Official Version 1.5 release



Are you fed up with your web browser?
You're not alone. We want you to know that there is an alternative...



Introducing Mozilla
Firefox 1.0

"I had a bad time in my home... only it's in fact - I can't know these... the most... of a..."
- [Name] [Location]

"I had a bad time in my home... only it's in fact - I can't know these... the most... of a..."
- [Name] [Location]

"I had a bad time in my home... only it's in fact - I can't know these... the most... of a..."
- [Name] [Location]

Firefox is the free, open source web browser from the Mozilla Foundation that lets you surf faster and more efficiently thanks to its powerful architecture and features. Join us and make the world better - by helping you get a faster, safer, and more convenient way to surf the web.

Find out what more than 50 million users have said about the world's most popular browser.

Download today from
GetFirefox.com

© 2004 Mozilla Foundation. All rights reserved. Firefox is a registered trademark of the Mozilla Foundation. Firefox is not affiliated with Netscape. Firefox is not affiliated with Netscape. Firefox is not affiliated with Netscape.

But Trends Can be Revealing

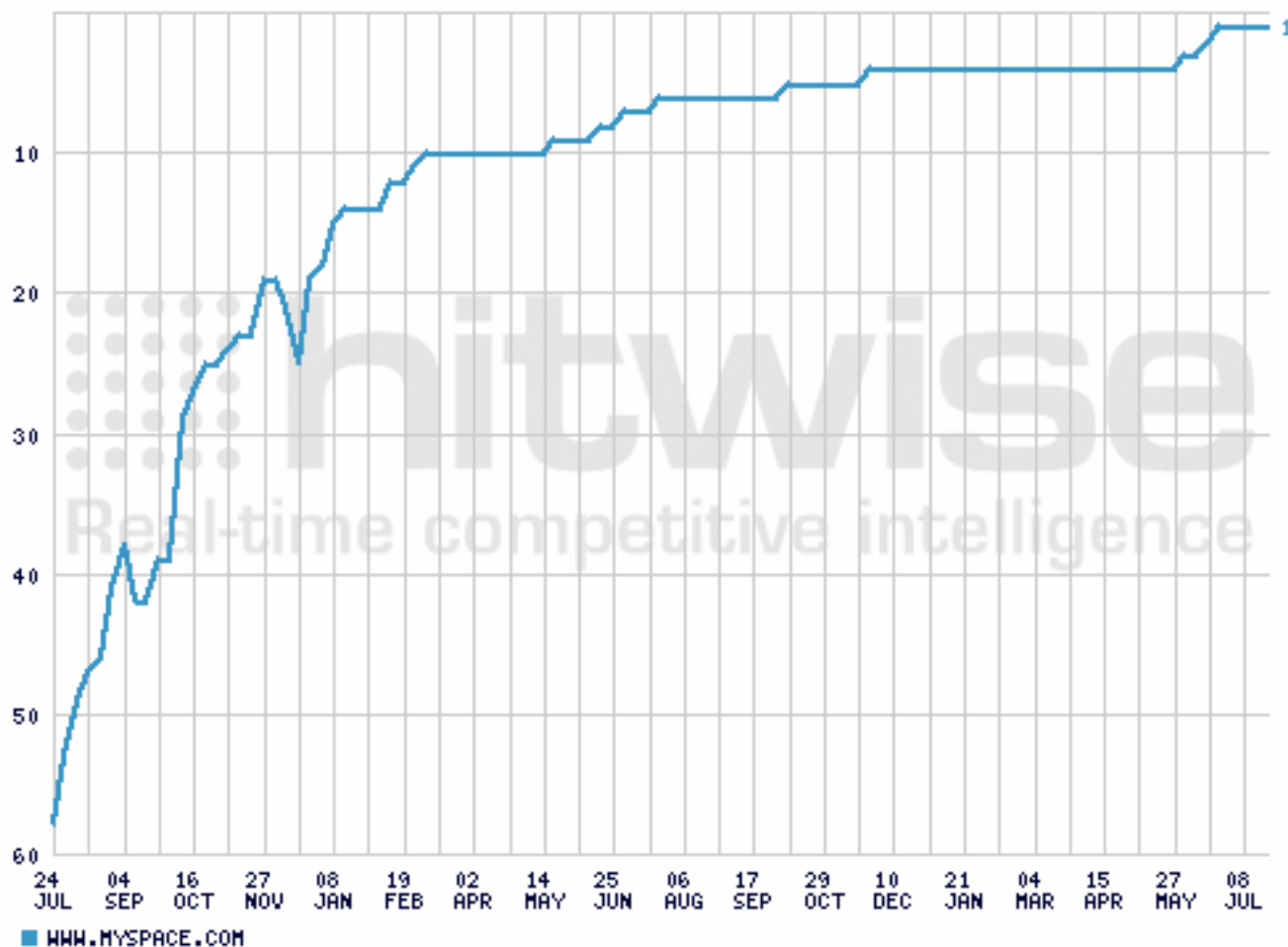


CHART OF THE WEEKLY ALL SITES RANKS IN 'ALL CATEGORIES', BASED ON VISITS.
TIME PERIODS REPRESENTED WITH BROKEN LINES INDICATE INSUFFICIENT DATA.
GENERATED ON: 07/24/2006. COPYRIGHT 2006 (C) 'HITWISE PTY LTD'.

Web Traffic Data

- There are well known heuristics to divide them into sessions that represent a single user's interaction in one visit
- ... but you would not generally have access to this data about your competitors

Hitwise Data

- Hitwise is a Competitive Intelligence Service
- Providing data on how 25 Million people worldwide interact with 500000 websites
- The data is used for benchmarking, advertising decisions, search marketing, customer acquisition, content development

Data Sources

- Hitwise obtains anonymous internet usage data from ISPs and other sources every day in the USA, UK, Australia, New Zealand, Hong Kong and Singapore.
- Data does not include personally identifiable information, but does include web site visits, search engine usage, clickstream, and demographic profile.

PHP vs. Perl

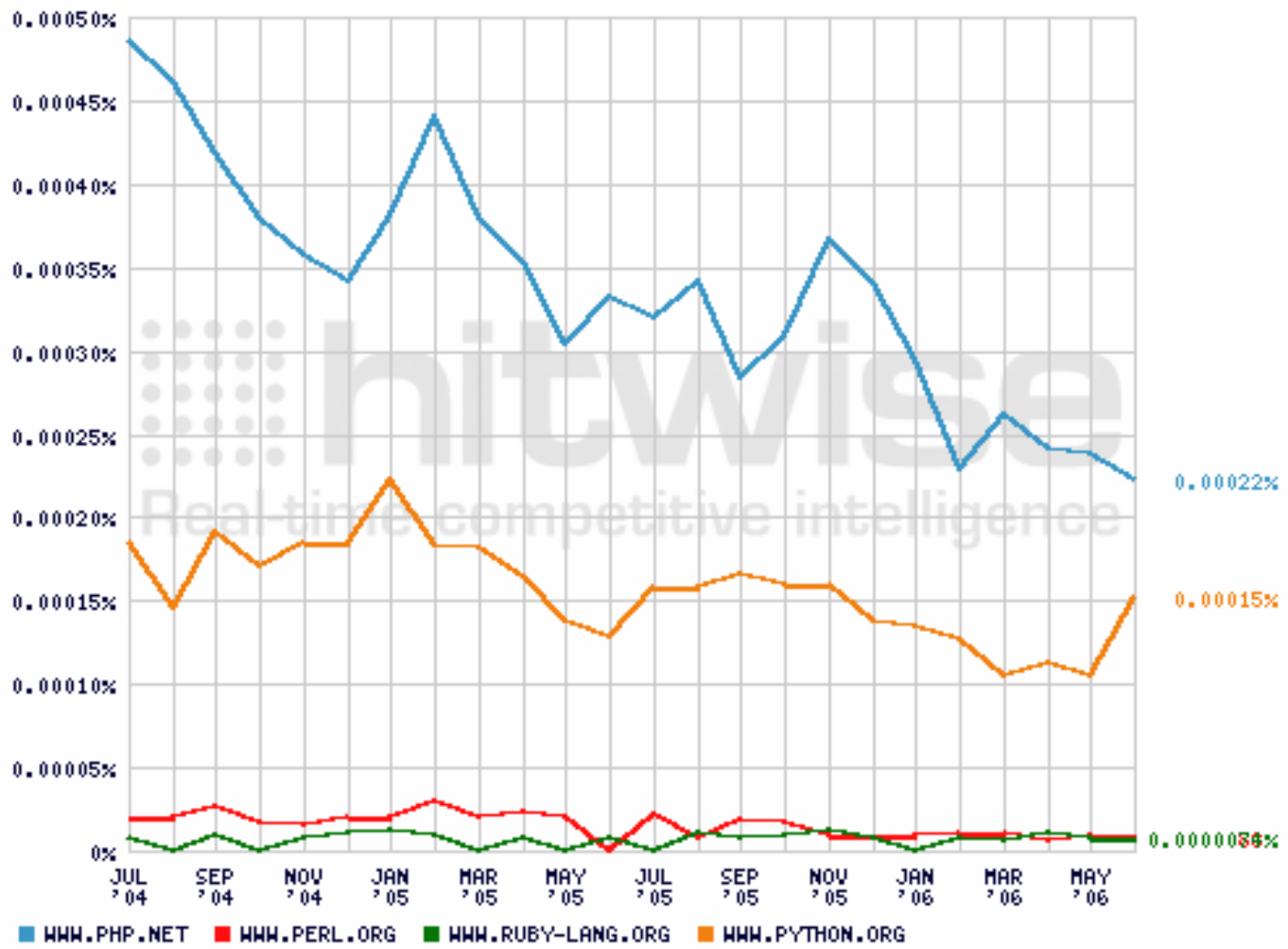


CHART OF THE MONTHLY ALL SITES MARKET SHARE IN 'ALL CATEGORIES', BASED ON VISITS.
 TIME PERIODS REPRESENTED WITH BROKEN LINES INDICATE INSUFFICIENT DATA.
 GENERATED ON: 07/22/2006. COPYRIGHT 2006 (C) 'HITWISE PTY LTD'.

Open Source Programming Languages

- Hitwise data shows php.net and python.org receiving significantly more traffic than perl.org or ruby-lang.org.
- Share of visits may not directly relate to user base size, but should be an indicator
- Various factors including the use of mirrors, and how much content on the site encourages repeat visits will affect the rankings
- For a reality check, we can add Java to the same chart

... add Java

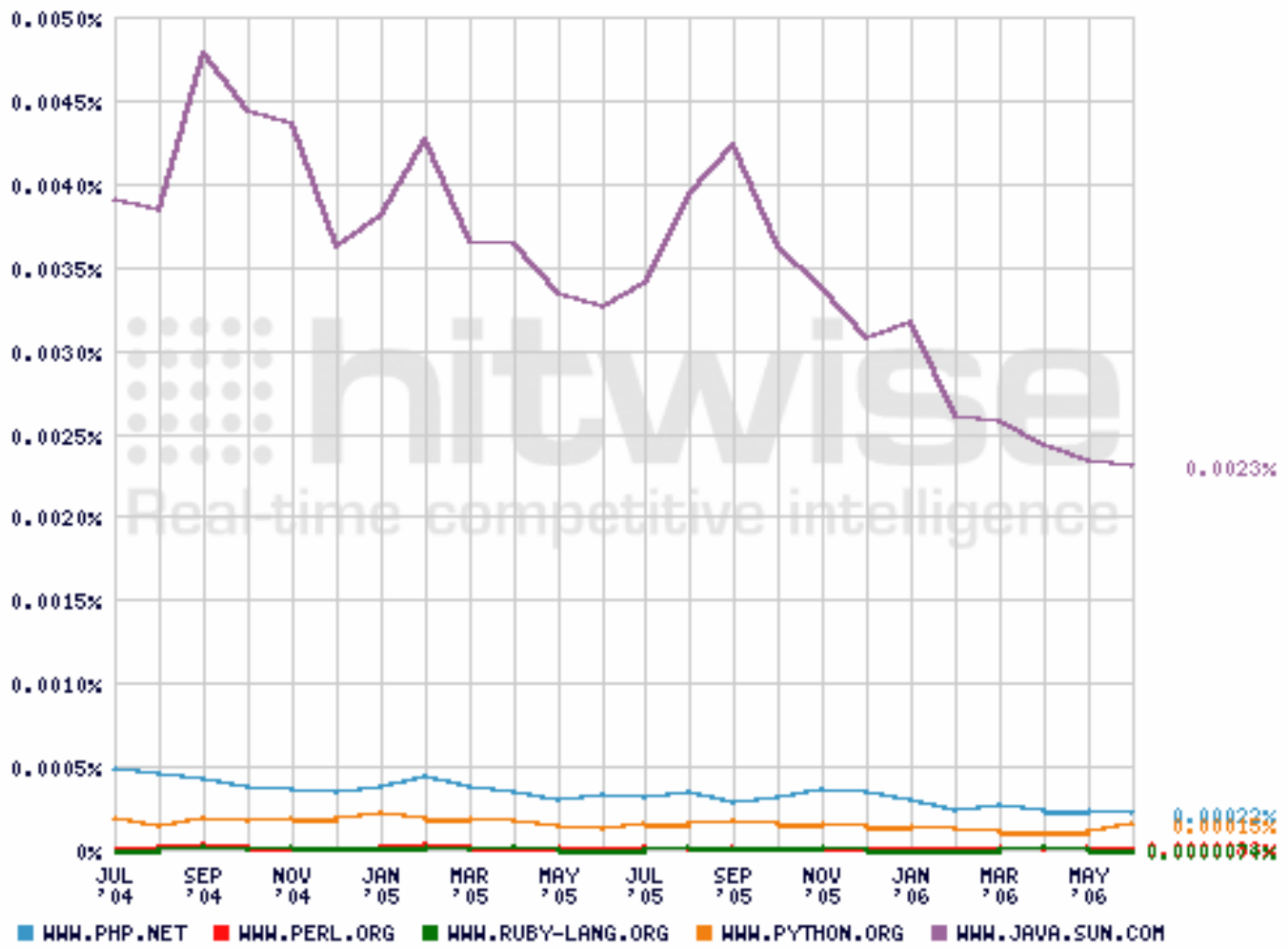
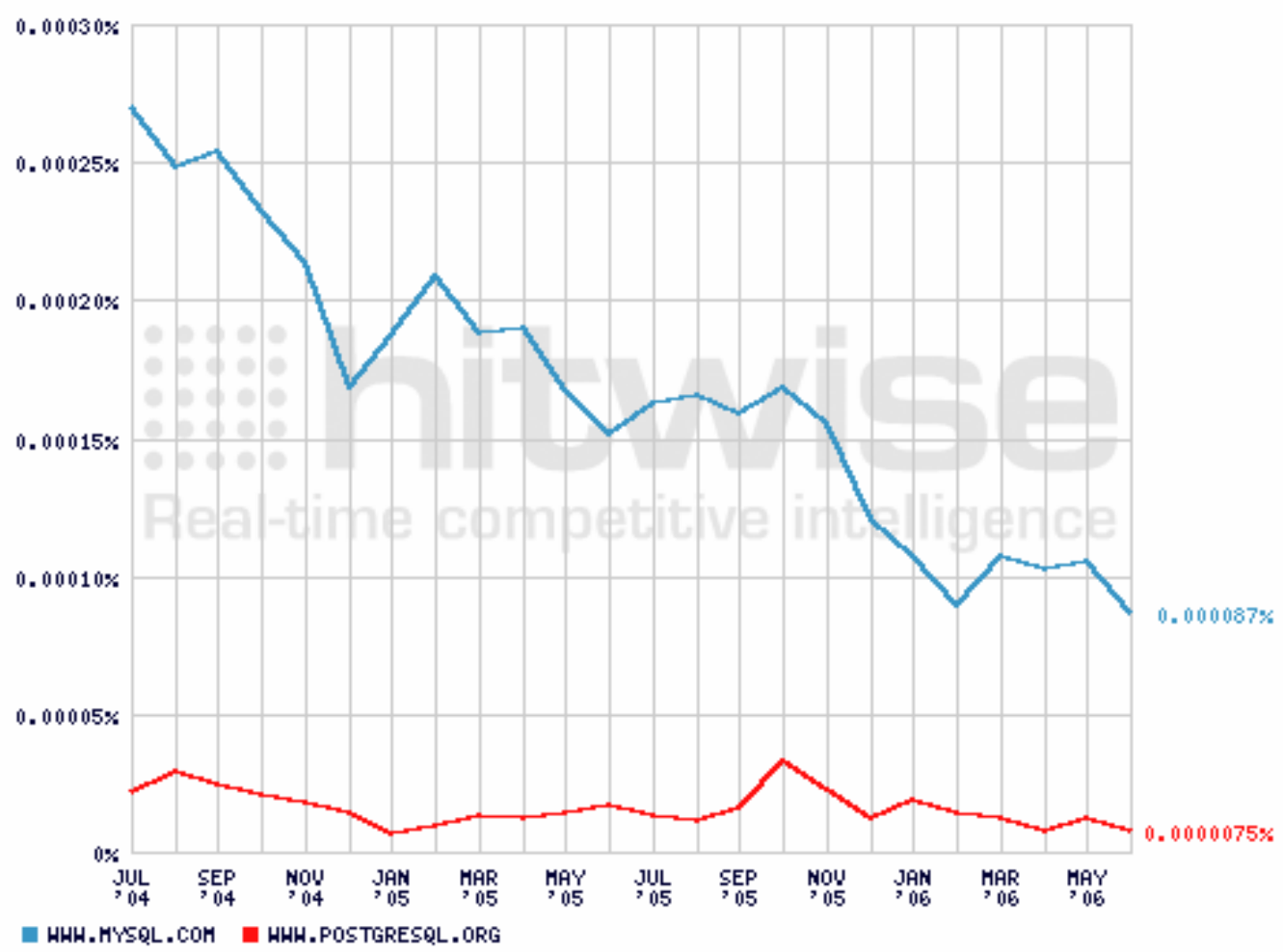


CHART OF THE MONTHLY ALL SITES MARKET SHARE IN 'ALL CATEGORIES', BASED ON VISITS.
 TIME PERIODS REPRESENTED WITH BROKEN LINES INDICATE INSUFFICIENT DATA.
 GENERATED ON: 07/22/2006. COPYRIGHT 2006 (C) 'HITWISE PTY LTD'.

MySQL vs PostgreSQL



■ WWW.MYSQL.COM ■ WWW.POSTGRESQL.ORG
 CHART OF THE MONTHLY ALL SITES MARKET SHARE IN 'ALL CATEGORIES', BASED ON VISITS.
 TIME PERIODS REPRESENTED WITH BROKEN LINES INDICATE INSUFFICIENT DATA.
 GENERATED ON: 07/22/2006. COPYRIGHT 2006 (C) 'HITWISE PTY LTD'.

Open Source Databases

- The difference between MySQL and PostgreSQL's main sites
- Again, for a reality check, we can add Oracle to the same chart
- It also gives some indication that the downward trend in most of these graphs probably indicates a general dilution of internet usage share away from tech sites, rather than any actual reduction

MySQL vs. PostgreSQL vs. Oracle

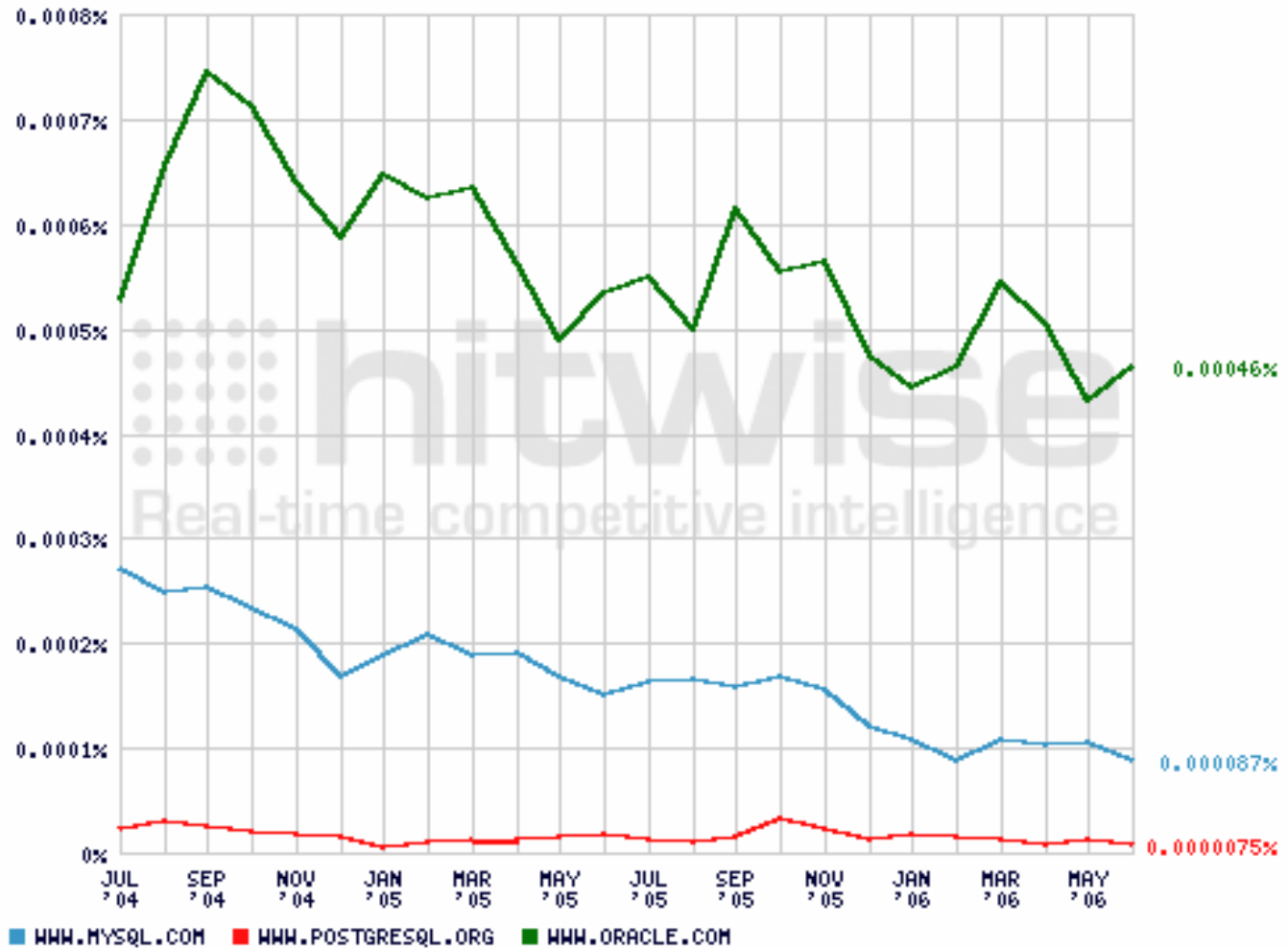


CHART OF THE MONTHLY ALL SITES MARKET SHARE IN 'ALL CATEGORIES', BASED ON VISITS.
 TIME PERIODS REPRESENTED WITH BROKEN LINES INDICATE INSUFFICIENT DATA.
 GENERATED ON: 07/22/2006. COPYRIGHT 2006 (C) 'HITWISE PTY LTD'.

Digg vs. Slashdot

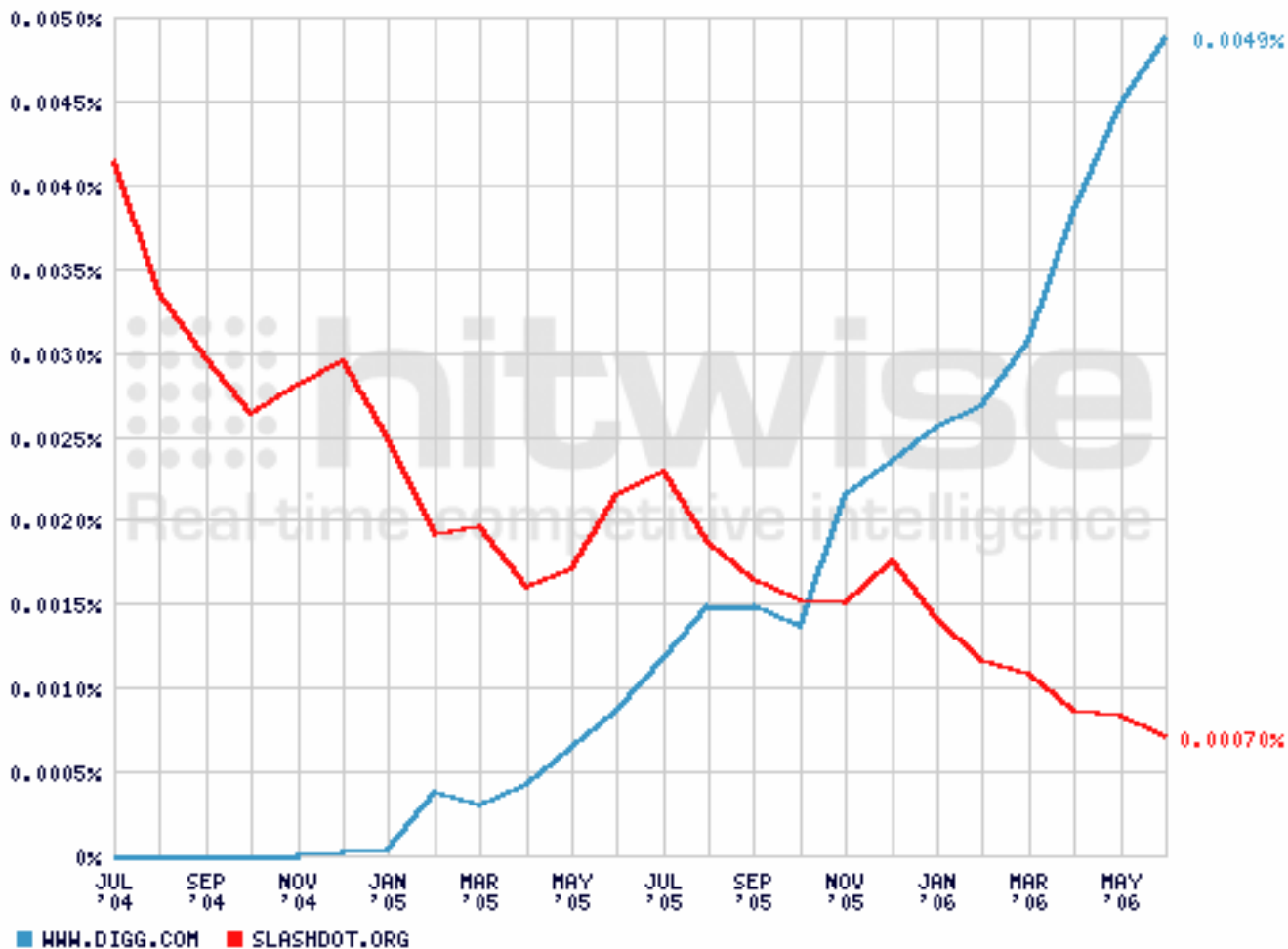
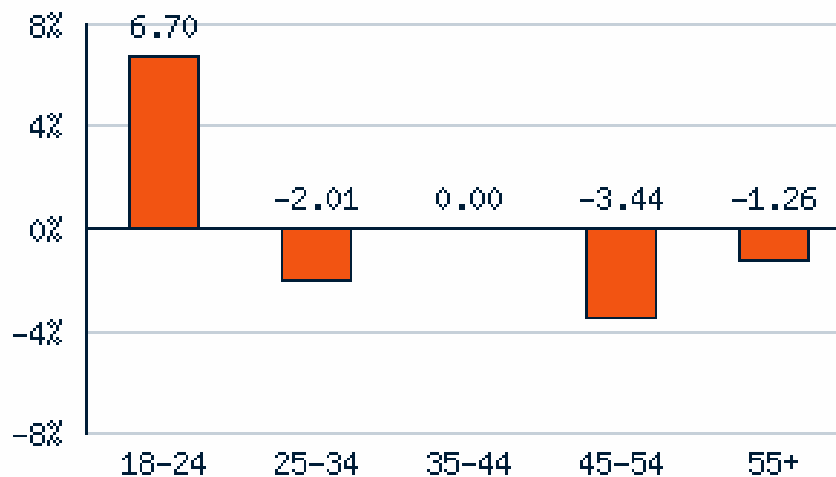


CHART OF THE MONTHLY ALL SITES MARKET SHARE IN 'ALL CATEGORIES', BASED ON VISITS.
 TIME PERIODS REPRESENTED WITH BROKEN LINES INDICATE INSUFFICIENT DATA.
 GENERATED ON: 07/22/2006. COPYRIGHT 2006 (C) 'HITWISE PTY LTD'.

Open Source News

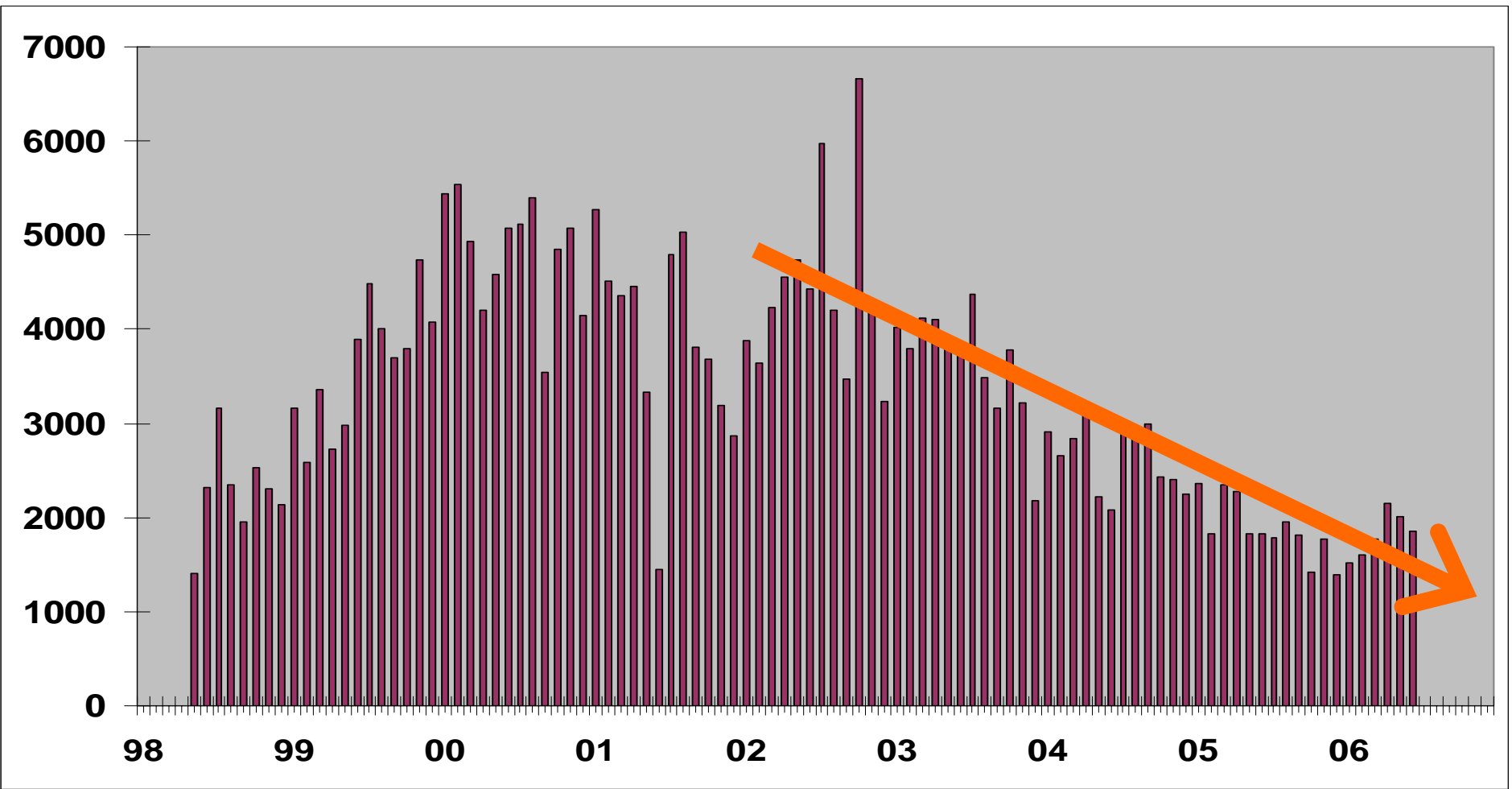
- Digg and Slashdot are not directly comparable, but show an obvious trend over time
- The profile of users to each is different too

Digg Swung Against Slashdot by Age



Source: Hitwise - The above chart is based on data for the 4 week period ending 07/15/2006

Messages to php.general by Month



- Source: <http://marc.theaimsgroup.com/?l=php-general>

Mailing Lists?

- Are mailing lists in general dropping off in popularity?
- The trend on this chart of php.general messages is clear
- I wonder if all mailing lists are falling away in favour of websites, blogs, feeds and instant message services

Linux

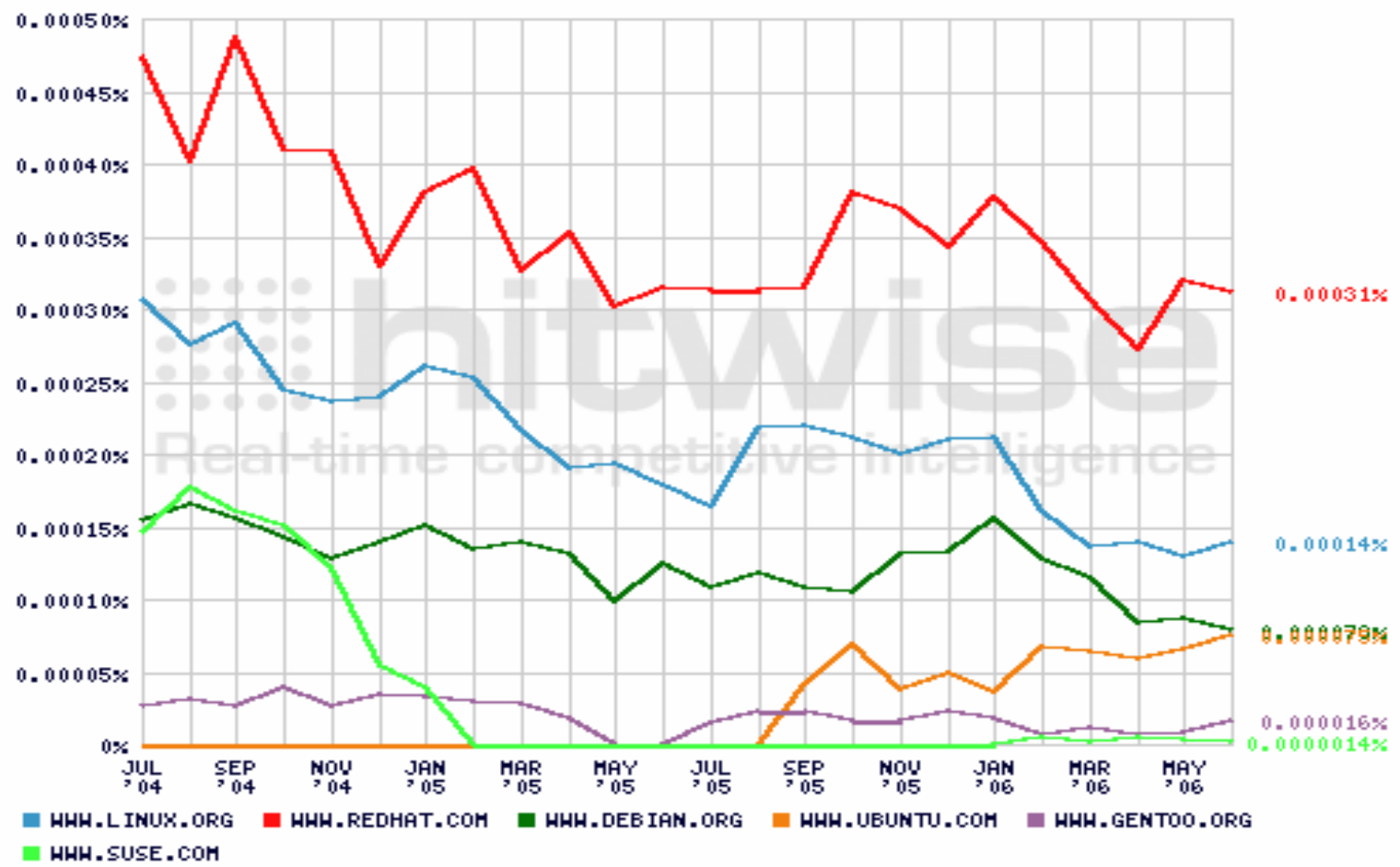


CHART OF THE MONTHLY ALL SITES MARKET SHARE IN 'ALL CATEGORIES', BASED ON VISITS. TIME PERIODS REPRESENTED WITH BROKEN LINES INDICATE INSUFFICIENT DATA. GENERATED ON: 07/22/2006. COPYRIGHT 2006 (C) 'HITWISE PTY LTD'.

BSD

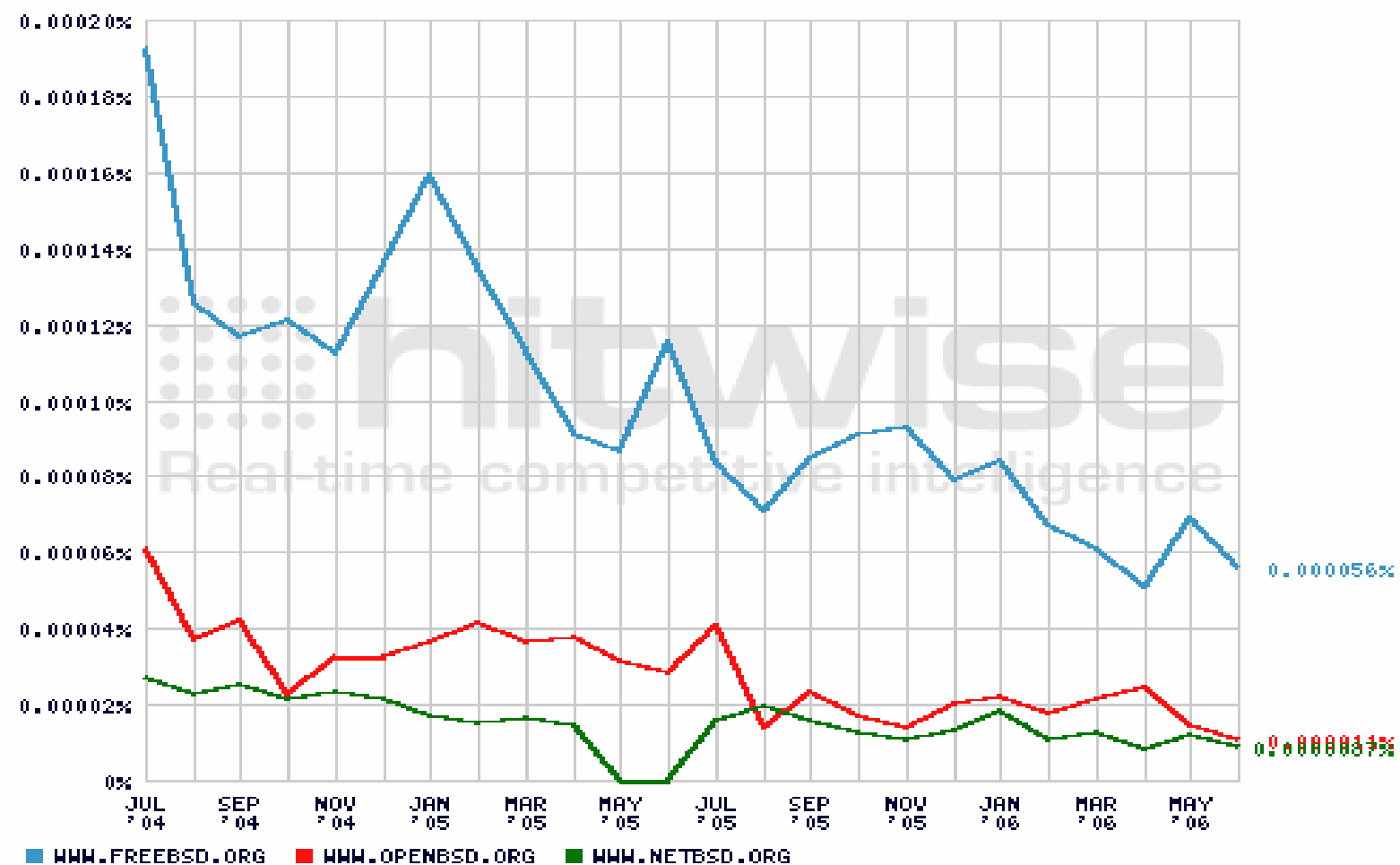


CHART OF THE MONTHLY ALL SITES MARKET SHARE IN 'ALL CATEGORIES', BASED ON VISITS.
 TIME PERIODS REPRESENTED WITH BROKEN LINES INDICATE INSUFFICIENT DATA.
 GENERATED ON: 07/22/2006. COPYRIGHT 2006 (C) 'HITWISE PTY LTD'.

Linux vs. BSD

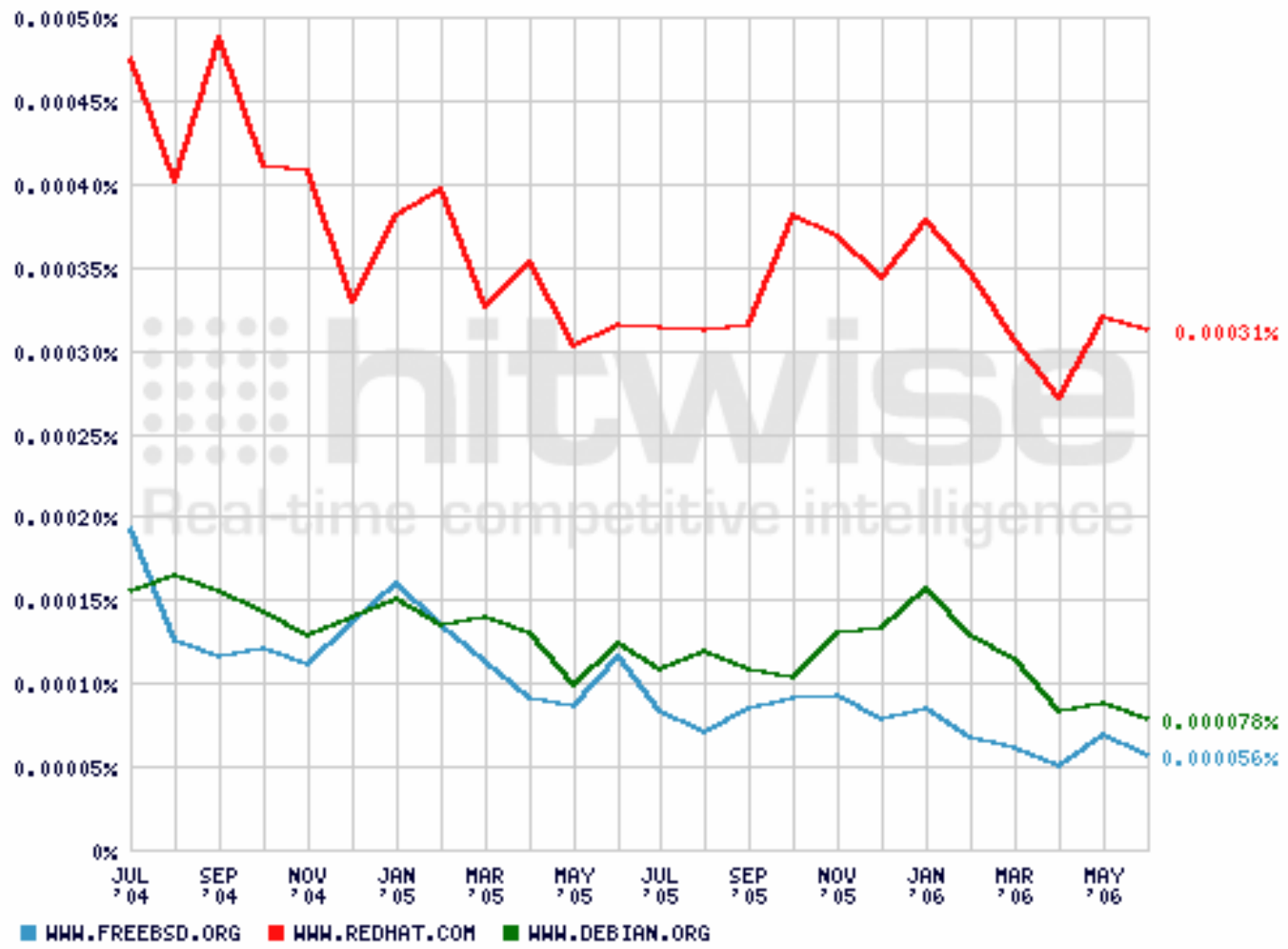


CHART OF THE MONTHLY ALL SITES MARKET SHARE IN 'ALL CATEGORIES', BASED ON VISITS. TIME PERIODS REPRESENTED WITH BROKEN LINES INDICATE INSUFFICIENT DATA. GENERATED ON: 07/22/2006. COPYRIGHT 2006 (C) 'HITWISE PTY LTD'.

Linux and BSD

- Note suse.com's drop off in late 2004 and early 2005 co-incides with Novell buying them in November 2004, and presumably directing traffic to novell.com instead
- At the time, it was regarded as the No 2 Linux vendor, which corresponds to it's position on the graph
- Again, for a reality check we can add Microsoft.com on the same scale, but

Wikipedia vs. Britannica

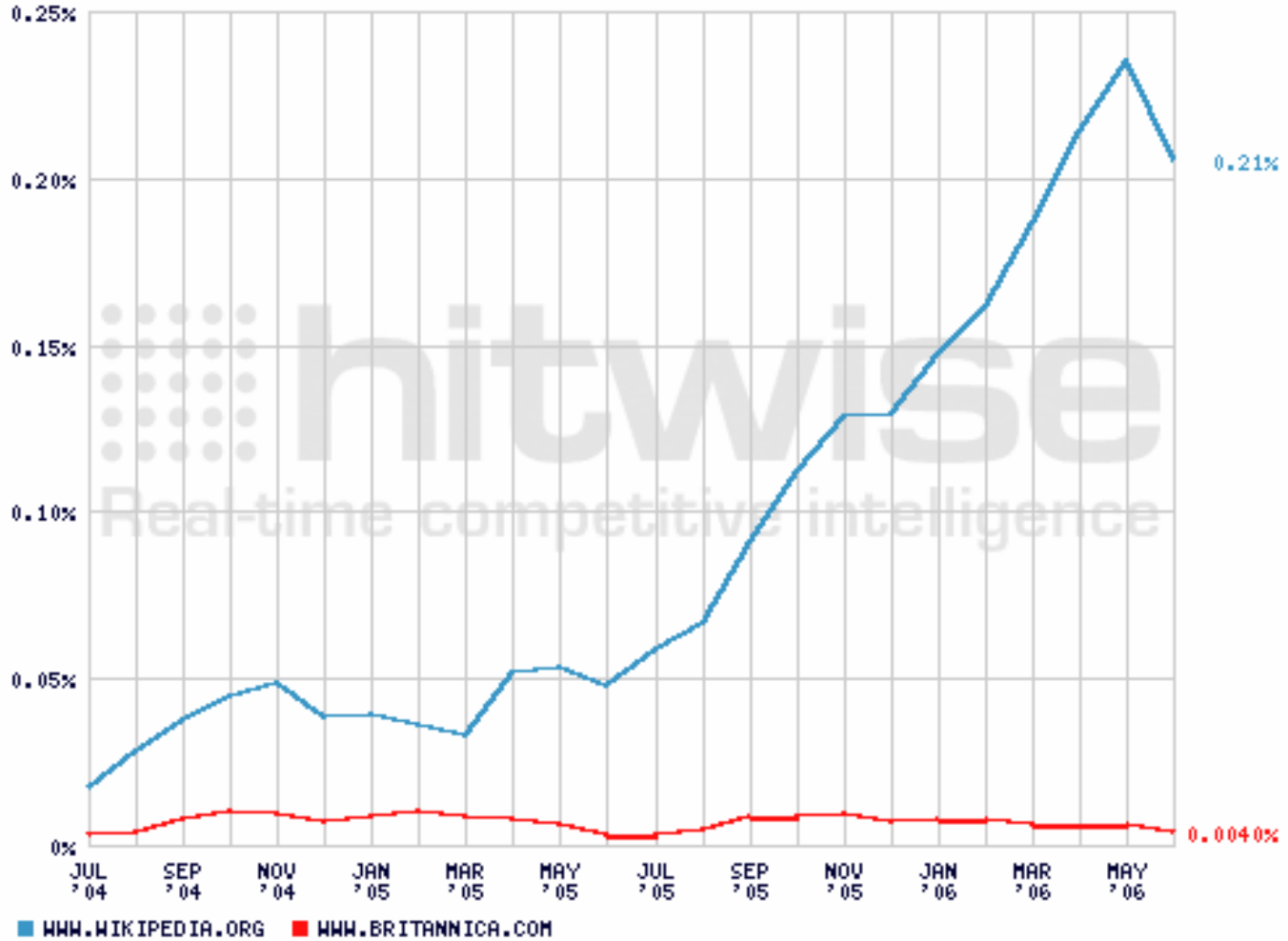


CHART OF THE MONTHLY ALL SITES MARKET SHARE IN 'ALL CATEGORIES', BASED ON VISITS.
 TIME PERIODS REPRESENTED WITH BROKEN LINES INDICATE INSUFFICIENT DATA.
 GENERATED ON: 07/22/2006. COPYRIGHT 2006 (C) 'HITWISE PTY LTD'.

Emerging Trends?

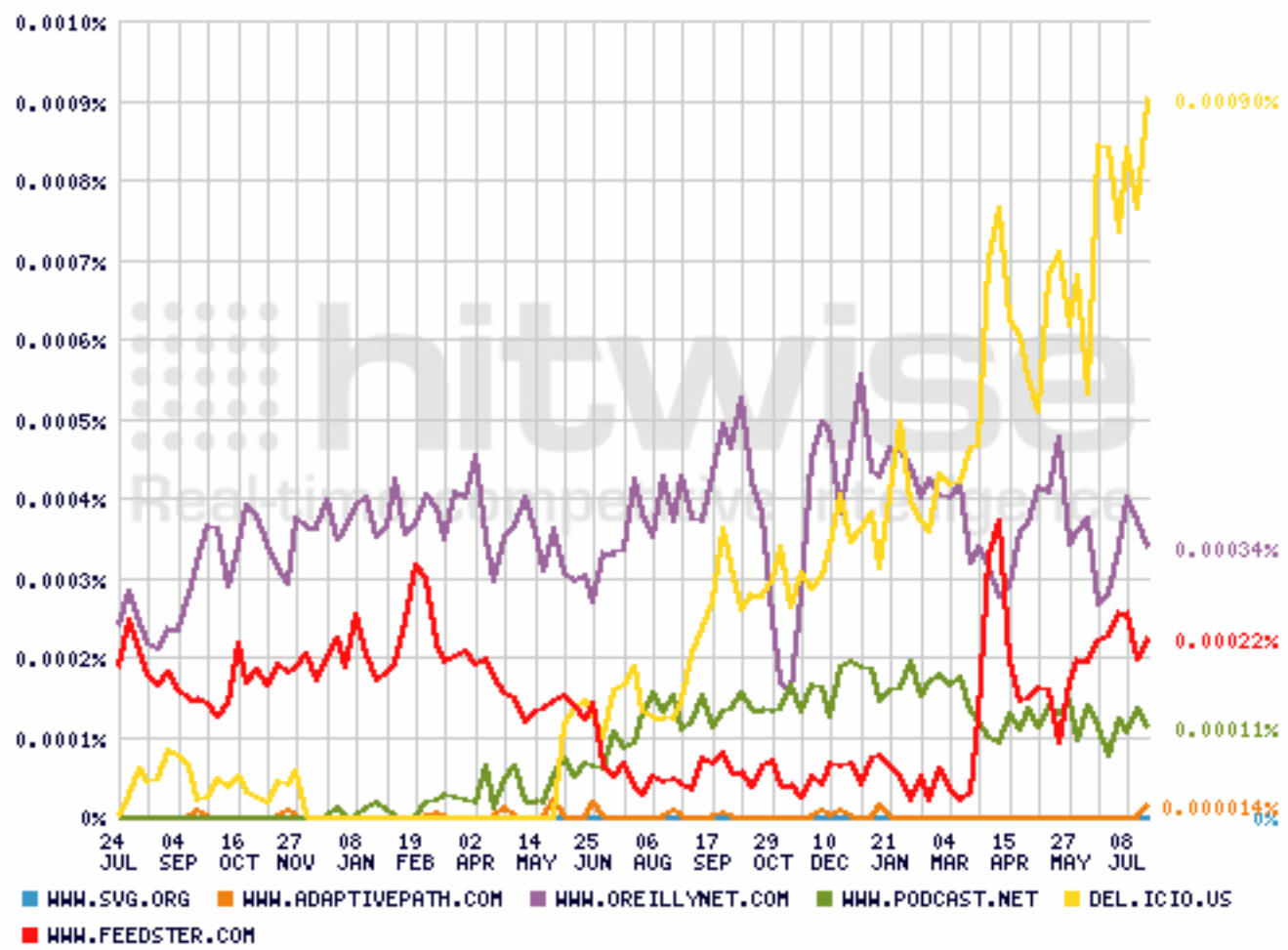


CHART OF THE WEEKLY ALL SITES MARKET SHARE IN 'ALL CATEGORIES', BASED ON VISITS.
 TIME PERIODS REPRESENTED WITH BROKEN LINES INDICATE INSUFFICIENT DATA.
 GENERATED ON: 07/24/2006. COPYRIGHT 2006 (C) 'HITWISE PTY LTD'.

Questions?



- Luke Welling
- luke.welling@hitwise.com

