

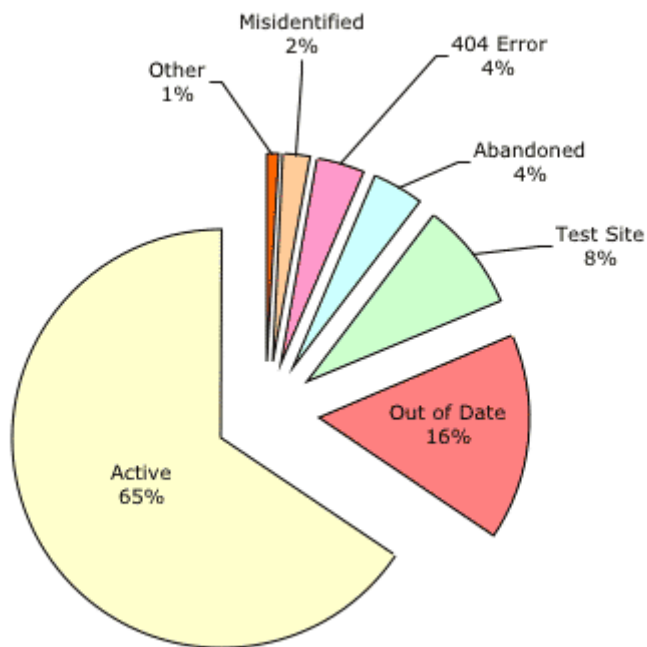
07.29.03

There has been some debate lately about the proper definition of "active" weblog.

The Blog Census gets its data from a variety of sources, including popular update sites and its own crawl of found links (described in euthanizing detail on the Methodology page). I was curious to see how accurate the census was in picking out active sites, and what kind of meaningful threshold we should be using to determine 'active' vs. 'out-of-date' weblogs. So on July 28, we picked a random sample of **529** weblogs from the census database, and examined them by hand.

To make sure the sample was truly random, we used radioactive decay data from the wonderful [hotbits](#) site - accept no pseudo-random substitutes!

Here's how our sample broke down:



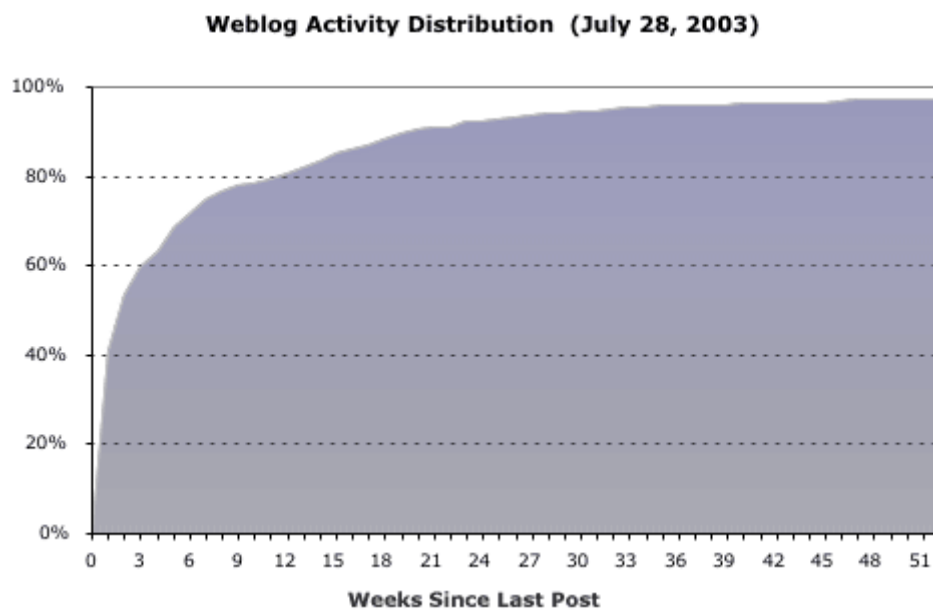
An explanation of the categories:

- **Active:** At least one post within the past eight weeks
- **Out-of-Date:** All posts older than eight weeks
- **Test Site:** A blog containing only test posts
- **Abandoned:** Blogs older than eight weeks where the author states that (s)he is through with blogging
- **404 Error:** Page not found
- **Misidentified:** Not a weblog
- **Other:** Password protected, suspended by administrator, etc.

This data suggests that about one in three weblogs in the census database is abandoned, unused, or very much out of date.

Cameron Marlow [points out](#) that there is no clear definition for what should constitute an "active" weblog. The threshold of eight weeks is completely arbitrary, so I thought it would be interesting to see the distribution of most recent posts over time.

The figure below shows the percentage of all actual weblogs from our sample (that is, the "active" and "out-of-date" slices of our pie chart) plotted against the number of weeks we want to use for a definition of "active". For example, 77% of the blogs in the sample had been posted to within the past eight weeks.



As you can see, the plot tails off at about the 95% level. Several blogs in the sample were abandoned as early as 2001.