



metaparadigm

Where the Developers Are: Open Source Activity Study

Michael Cark <michael at metaparadigm dot com>

Scott McNeil <mcneil at globaldesktop dot com>

<http://oss.metaparadigm.com/oss-activity/>

Presentation Overview

• Goal

- Determine a fair measure of open source software contribution per country.

• Method

- Analyze activity on public Internet mailing lists of major open source projects.

Goals

- **Global picture of OSS activity.**
- **Fair**
 - Don't favor one country or geographic location.
 - Don't favor infrastructure rich countries.
- **Measure activity data from a wide range of open source projects.**
- **Beginning of a larger project to gauge OSS 'Use vs. Contribution'.**

Methodology

- **Use open source development mailing list posts as a measure of activity**
 - 2 years of mailing list data mirrored from gmane was studied from Jan 2003 through Dec 2004
- **Selected major open source projects**
 - Apache, Debian, Fedora, Gnome, JBoss, KDE, Linux kernel, OpenLDAP, Samba, Sendmail, X Windows
- **Use per Internet Capita as a normative baseline (US Census and ITU data).**



Methodology (cont.)

- **Account email addresses to countries using TLD lookup and IP GEO location database**
 - Advantage, can account .com, .net and .org to foreign countries.
- **Remove SPAM**
 - MX verification.
- **Remove outliers that can't be accounted**
 - Gnome.org, redhat.org, kde.org, debian.org ...
 - Should not effect rankings as proportionally distributed.



Assumptions and Constraints

- **English language only**

- Most global OSS development in English.
- Need more research on normative measure for language specific communities!!!

- **Type of discourse**

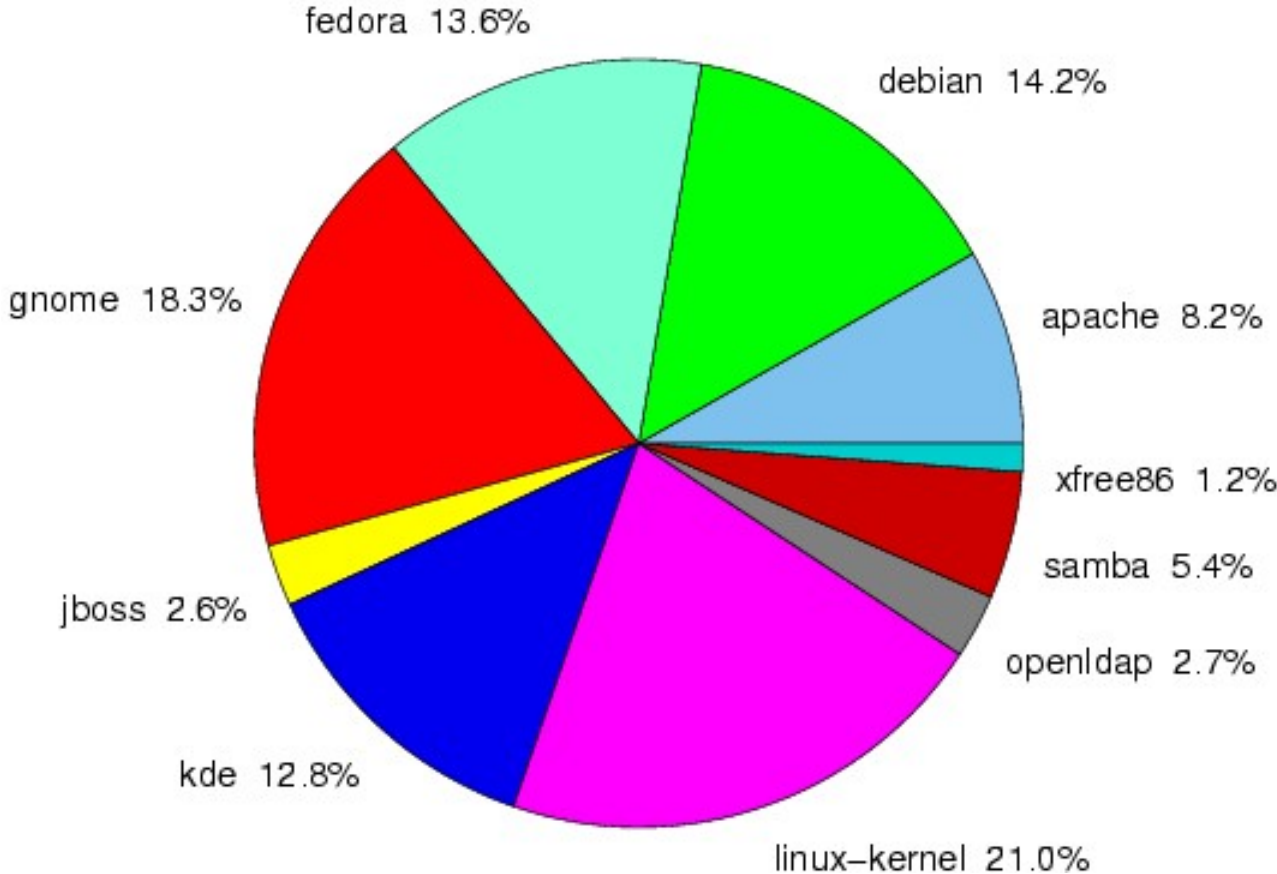
- Assumption that posts are related to development and contribution i.e. bugs, patches, etc.



Disclaimer

- **Version 0.5**
- **Presently unpublished research.**
- **Statistics on degree of error needs to be established.**
- **Need to broaden data set**
 - currently analyzed approximately 1 million emails.
- **Contact us to participate in next version**
 - Source code for analysis to be released.

Mailing list distribution



Processing Statistics

Group	Proc	Total	TLD	Geo	Excl.	Bad MX	No Geo	Done
gmane.comp.apache.apr.devel	3897	4594	866	3031	377	320	0	84.80%
gmane.comp.apache.devel	8642	10468	1848	6794	1088	738	0	82.60%
gmane.comp.apache.documentation	1699	1974	1025	674	213	62	0	86.10%
gmane.comp.apache.incubator.general	3140	4669	564	2576	1477	52	0	67.30%
gmane.comp.apache.mod-perl	12544	13495	2387	10157	13	937	1	93.00%
gmane.comp.apache.mod-perl.devel	5257	6326	353	4904	10	1059	0	83.10%
gmane.comp.apache.mod-proxy	344	358	134	210	0	14	0	96.10%
gmane.comp.apache.mod-ruby	347	357	60	287	1	9	0	97.20%
gmane.comp.apache.mod-security.user	384	394	68	316	0	10	0	97.50%
gmane.comp.apache.mod-ssl.user	1308	1430	381	927	41	80	1	91.50%
gmane.comp.apache.user	27924	29249	8385	19539	112	1209	4	95.50%
gmane.comp.gnome.accessibility.devel	792	823	170	622	24	7	0	96.20%
gmane.comp.gnome.accessibility.general	2479	2654	409	2070	49	126	0	93.40%
gmane.comp.gnome.apps.balsa	3172	3472	1989	1183	170	130	0	91.40%
gmane.comp.gnome.apps.epiphany	2713	3268	1115	1598	404	150	1	83.00%
gmane.comp.gnome.apps.gnomemeeting	4369	6462	2316	2053	19	2074	0	67.60%
gmane.comp.gnome.apps.gnucash.devel	4590	4824	1399	3191	51	183	0	95.10%
<hr/>								
gmane.linux.kernel	167333	179450	59778	107555	3015	9078	24	93.20%
gmane.linux.redhat.fedora.devel	17926	18913	3844	14082	21	966	0	94.80%
gmane.linux.redhat.fedora.general	90543	96699	27260	63283	344	5489	323	93.60%
gmane.network.openldap.devel	2256	2297	636	1620	25	16	0	98.20%
gmane.network.openldap.general	19277	19874	7411	11866	28	522	47	97.00%
gmane.network.samba.general	34187	41832	13666	20521	4853	2536	256	81.70%
gmane.network.samba.internals	8491	12430	3083	5408	3682	257	0	68.30%
gmane.os.freebsd.devel.gnome	11294	11502	1903	9391	8	195	5	98.20%
all	795748	990762	289355	506393	151704	42529	781	80.30%



Top 24

Rank	Posts / K Users	Country	Unique Posts	Unique Addr	Unique Hosts	Internet Pen. %	Internet Pop. (M)
1	3.48	Norway	5506	611	360	34.6	1.6
2	2.94	Czech Republic	8079	541	290	26.8	2.7
3	2.69	Switzerland	7022	887	537	35.1	2.6
4	2.55	Germany	99184	8437	4244	47.3	39
5	2.51	Netherlands	21345	1643	940	52.2	8.5
6	2.48	Ireland	3086	192	121	31.3	1.2
7	2.38	Sweden	12258	1085	594	57.3	5.1
8	2.35	Denmark	6523	615	467	51.3	2.8
9	2.33	United States	376282	37644	15612	55.1	161.6
10	2.32	Austria	8766	776	458	46.2	3.8
11	2.14	Belgium	7275	699	346	32.8	3.4
12	2.03	Australia	22854	2029	1160	56.7	11.3
13	1.92	United Kingdom	48972	4438	2792	42.3	25.5
14	1.91	Hungary	4446	406	247	23.2	2.3
15	1.9	Finland	5052	636	293	50.9	2.7
16	1.63	France	36040	3374	1472	36.6	22.1
17	1.59	Portugal	3231	347	168	19.4	2
18	1.53	Canada	25548	2902	1629	51.3	16.7
19	1.35	New Zealand	2837	644	459	52.6	2.1
20	1.31	Israel	2442	299	174	30.1	1.9
21	1.1	South Africa	3205	424	308	6.8	2.9
22	0.99	Russia	5834	1039	536	4.1	5.9
	0.95	<i>All countries</i>	795748	81897	39932	13.5	839.4
23	0.91	Spain	8779	1261	603	23.9	9.6
24	0.8	Poland	7144	783	440	23.2	9



Top 25 - 49

Rank	Posts / K Users	Country	Unique Posts	Unique Adrrs	Unique Hosts	Internet Pen. %	Internet Pop. (M)
25	0.7	Bulgaria	1088	117	87	20.6	1.5
26	0.66	Italy	12843	1779	781	33.7	19.5
27	0.57	Brazil	8603	1239	665	8.2	15.1
28	0.48	Slovakia	661	145	109	25.6	1.4
29	0.46	Greece	738	125	75	15	1.6
30	0.41	Hong Kong	1319	365	171	47.2	3.2
31	0.38	Romania	1605	248	160	19.1	4.3
32	0.3	Argentina	1323	309	190	11.2	4.4
33	0.27	Venezuela	345	44	24	5.1	1.3
34	0.2	India	3773	974	339	1.7	18.6
35	0.2	Turkey	1117	167	118	8.1	5.5
36	0.2	Chile	864	120	84	27.2	4.3
37	0.18	Philippines	673	82	65	4.4	3.8
38	0.17	Singapore	383	117	70	50.9	2.2
39	0.14	Belarus	207	26	11	14.1	1.5
40	0.09	Japan	5730	653	465	48.3	61.5
41	0.07	Taiwan	622	202	136	39.1	8.9
42	0.07	Indonesia	592	97	74	3.8	9
43	0.06	Mexico	703	147	103	11.8	12.4
44	0.06	Malaysia	454	85	66	34.4	8.1
45	0.06	Colombia	145	40	34	6.2	2.6
46	0.03	Vietnam	105	21	17	4.3	3.6
47	0.03	South Korea	852	236	92	61	29.6
48	0.02	China	1787	725	260	6.3	82.1
49	0.02	Thailand	123	40	28	11.1	7.2



Conclusion

- **Western Europe contributes the most.**
 - This is almost all in a second language.
- **Needs further work for study of non-English language development.**
- **Looking for contributors to this study!**
- **Results are on the web here:**
 - <http://oss.metaparadigm.com/oss-activity/>

Thank You

Michael Clark <michael at metaparadigm dot com>

Scott McNeil <mcneil at globaldesktop dot com>

<http://oss.metaparadigm.com/oss-activity/>