

The Lives of Others

Open-Source Development Practices Elsewhere

Peter Eisentraut

Senior Software Engineer
Lab Development
F-Secure Corporation

PGCon 2009



Ground Rules



- ▶ I don't know everything.
- ▶ If you are involved in other projects, chime in.
- ▶ Discussion welcome

Whom To Learn From?

- ▶ Look at the biggest open-source projects
- ▶ Metric: installation footprint, installation count
- ▶ Where to find those numbers?
 - ▶ the Universal Debian Database
 - ▶ doesn't count OS distributions such as Debian, Fedora, FreeBSD, OpenSolaris
 - ▶ Java also underrepresented



Sidetrack: The Universal Debian Database

Example: query for packages with largest install footprint and popularity

```
SELECT rank() OVER (ORDER BY score DESC),
       source,
       sum(installed_size::numeric * insts) AS score
FROM packages JOIN popcon_src USING (source)
WHERE distribution = 'debian'
      AND release = 'lenny'
      AND component = 'main'
      AND architecture IN ('all', 'i386')
GROUP BY source
ORDER BY score DESC
LIMIT 100;
```

The Biggest Projects

OpenOffice.org, Linux, GCC, GCJ, Qt, Mesa, KDE, Glibc, Firefox, MySQL, Boost, Perl, Samba, LDP, Vim, Gtk+, FFmpeg, OPAL, Evolution, OpenJDK, Gtk—, wxWidgets, PyQt, Gimp, TeX Live, X.org, SCOWL, OpenSSL, Thunderbird, Xalan, Berkeley DB, ncurses, Foomatic, SELinux, Gutenprint, PHP, GNU binutils, Python, KOffice, Wine, Ruby, ICU, LAPACK, Webkit, PostgreSQL, glib, libsig++, Inkscape, PWLib, GNOME, Apache, HPLIP, Xerces, Libxml, GNU Gettext, CUPS, SANE, Net-SNMP, Ghostscript, Mono, Eclipse, Pidgin

PostgreSQL — A Role Model?

Things to keep in mind:

1. Process details are hard to gather for outsiders.
2. Other projects are not without problems.
3. Many of them have worse problems than we do.
4. Others look at the PostgreSQL project for advice and inspiration.
5. It's amazing that anything gets done. :-)

Version Control

- ▶ In the lead: Git, Subversion
- ▶ Also-rans: CVS, Mercurial, Bazaar
- ▶ Commercial systems phased out
- ▶ Automatic conversion services in use

Build Tools

- ▶ Autotools continue to rule!
- ▶ CMake might be challenger
 - ▶ well-known usage: KDE, MySQL (for Windows only)
 - ▶ uptake still slow (88 of 12000 packages in Debian)
- ▶ That's it, for C programs.

Documentation Tools

- ▶ DocBook emerged as the standard.
 - ▶ XML vs. SGML
- ▶ Texinfo (only at GNU)
- ▶ man pages
- ▶ HTML
- ▶ some specialized solutions:
 - ▶ Python: reStructuredText
 - ▶ Perl: Perldoc
 - ▶ Qt: qdoc

Discussion Forums

- ▶ Mailing lists are clearly the standard.
- ▶ But don't argue about Reply-To!
- ▶ Usenet is dead. :-)
- ▶ Web-forums serve support requests and newbies better.
- ▶ ...and then there are bug-tracking systems.

Bug Tracking

- ▶ Bugzilla is king (often heavily themed and customized).
- ▶ contenders: Roundup, Jira, *forge
- ▶ marginal: RT

Coding Styles

- ▶ Coding style appears to be an issue specific to C/C++.
- ▶ Besides the GNU style, various random styles can be found.
- ▶ CVS \$Keywords\$ are pretty much gone.
- ▶ Patch styles (-c vs. -u) are an unsettled issue.

- ▶ About 1/3 of projects have active wikis.
- ▶ MediaWiki is most popular.

The Foundries

- ▶ Project-specific *forge services are rare.
- ▶ More popular: Common repo and Bugzilla for everyone.
- ▶ This has implications for management of sub-/sister projects.



Write Your Own Tools

Some tools are side-products of other projects:

- ▶ ccache
- ▶ tmalloc
- ▶ distcc
- ▶ Bugzilla
- ▶ PatchTracker
- ▶ Contributions to build tools (autoconf, cmake, etc.) are welcome.

Release Scheduling

Trend is toward time-based releases:

- ▶ OpenBSD (6 months)
- ▶ GNOME (6 months)
- ▶ Ubuntu (6 months)
- ▶ Debian (18 months)
- ▶ OpenOffice.org (6 months)
- ▶ PostgreSQL (12 months)
- ▶ Linux kernel (3 months)
- ▶ (MySQL)
- ▶ ...

Note: The upgrade issue is fairly unique to PostgreSQL.

Minor Releases

- ▶ Stable + development branch is standard.
- ▶ Some maintain more than one stable, e.g., Samba.
- ▶ PostgreSQL's 5+ back branches have got to be a record.
- ▶ MySQL might come close.
- ▶ Minor releases can also have a fixed schedule.
(OpenOffice.org: 3 months)

The Rewrite Trap

Major reengineering efforts usually have schedule and/or quality problems. Recall:

- ▶ Samba 4
- ▶ Perl 6
- ▶ KDE 4

Pre-Releases

- ▶ Beta, RC are standard.
- ▶ Even/odd model (Linux, Perl) is obsolescent.
- ▶ Alpha/milestone releases are rare.

Release Manager

- ▶ Some projects appoint rotating release managers.
- ▶ In other projects, these roles fall into place by tradition.

Downstream Packaging

Various approaches:

- ▶ Tight integration, e.g.: PostgreSQL, KDE, Samba
- ▶ Laissez-faire, e.g.: Linux, GCC, MySQL
- ▶ Conflict-prone, e.g.: OpenOffice.org, Mozilla



Security Handling



- ▶ Most sizeable projects have special security contacts/teams.
- ▶ CVE numbers are standard.
- ▶ Note: Non-database projects have a skewed/different view of “security”.

Committers

- ▶ Surprisingly, many projects seem to do with very few committers.
- ▶ Some projects have very large amounts of committers.
- ▶ I think this is strongly related to the software architecture and project layout. Compare:
 - ▶ KDE has everything in one repo; hundreds of committers.
 - ▶ PostgreSQL has PgFoundry etc.
 - ▶ Linux kernel uses a very different model.
- ▶ Managing external committers tends to be a problem for company-run projects.

Project Steering

- ▶ Company-run projects have a clear direction (usually).
- ▶ In principle, the developers create their own destiny everywhere else.
- ▶ Some projects have elected boards etc., but these don't have technical influence.
- ▶ Most projects are steered by the “old farts”.

Forking

- ▶ Forking continues to happen.
- ▶ Contemporary cases: Go-OO.org, eglibc, OurDelta, Iceweasel
- ▶ Reasons:
 - ▶ Faster/different feature advancement
 - ▶ Licensing or trademark conflicts
 - ▶ Company vs. community barrier
 - ▶ Personal problems
- ▶ Many commercial forks of open-source DBMS:
EnterpriseDB, Greenplum, Truviso, Kickfire, Infobright

Licensing

- ▶ Distributors are incredibly picky.
- ▶ So are legal departments.
- ▶ Well-known licenses are important.
- ▶ Consistent and uniform licensing helps.
- ▶ Weird licensing can kill projects.
- ▶ Many projects have had their share of problems: KDE (Qt), Linux (SCO, firmware), XFree86, Mambo, Apache vs. OpenBSD, PostgreSQL (GPL), OpenSSL vs. Debian

Copyright Notices

- ▶ Amateur legal opinion: Copyright notices are more or less meaningless. But they could help if you plan to go to court.
- ▶ Most company-sponsored code has standard copyright headers.
- ▶ FSF and Apache use copyright assignments and careful changelog tracking.
- ▶ Most other projects are pretty random.
 - ▶ Initial author: Perl, Samba
 - ▶ Various authors: KDE, Vim
 - ▶ “Fake” organizations: OpenSSL, (PHP), PostgreSQL, wxWidgets
 - ▶ *nothing*: Python
- ▶ Many (smaller) projects apply the GPL wrongly.

Legal and Funds Management

Four models:

- ▶ Run and owned by a single company: OpenOffice.org, MySQL, Qt, Berkeley DB, CUPS, (PHP)
- ▶ Run and controlled by a non-profit association: GCC, KDE, Mozilla, GNOME, Apache, (Fedora)
- ▶ Loosely organized but someone in the background: Linux, PostgreSQL, Debian, OpenBSD
- ▶ Unorganized: Vim

Standards Organizations

OSS participation is possible, e. g.:

- ▶ OpenOffice.org
- ▶ Linux/glibc
- ▶ KDE/GNOME/X.org/Freedesktop
- ▶ MySQL, PostgreSQL, ... ?

The Average Sizeable Open-Source Project 2010/2011

- ▶ Written in C (or Java)
- ▶ GPL
- ▶ Source in Git
- ▶ Bugs in Bugzilla
- ▶ Built with Autoconf (or Maven)
- ▶ Released every 6 months

Open Issues for PostgreSQL

- ▶ Version control system
- ▶ Web forum
- ▶ Bug tracking system
- ▶ CMake?
- ▶ DocBook XML
- ▶ Release cycle length
- ▶ Upgradability



Rights and Attributions

This presentation “The Lives of Others: Open-Source Development Practices Elsewhere” was authored by Peter Eisentraut and is licensed under the [Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported license](#).

- ▶ The image on page 2 is “[Comiskey Park \(US Cellular Field\)](#)” by Flickr user Zol87, available under the [Creative Commons Attribution-Noncommercial 2.0 Generic license](#).
- ▶ The image on page 3 is “[Apples and oranges](#)” by Flickr user Jeff Tabaco, available under the [Creative Commons Attribution-Noncommercial-Share Alike 2.0 Generic license](#).
- ▶ The image on page 14 is “[Valimon asema Helsinki](#)” from Wikipedia and is in the public domain.
- ▶ The image on page 21 is “[Waterfalls, Mesa Arts Center](#)” by Flickr user cobalt123, available under the [Creative Commons Attribution-Noncommercial-Share Alike 2.0 Generic license](#).
- ▶ The image on page 22 is “[Suomenlinna](#)” by Flickr user Anosmia, available under the [Creative Commons Attribution 2.0 Generic license](#).
- ▶ The image on page 31 is “[Committee meeting](#)” by Flickr user Matt and Kim Rudge, available under the [Creative Commons Attribution-Noncommercial 2.0 Generic license](#).