π's in Debian or Scientific Debian: NumPy, SciPy and beyond

Yaroslav O. Halchenko

Debian Project,
Dartmouth College, USA

EuroScipy 2011, Paris, France

Scientists report:

[illegible] GNU/ Linux users report the lowest average time they need to invest in maintenance of their personal computing environment (5.77 h/ month). [illegible]

Hanke M and Halchenko YO (2011) *Neuroscience runs on GNU/Linux.* Front. Neuroinform. 5:8. doi: 10.3389/fninf.2011.00008

Debian: once upon a time

Fellow Linuxers,

This is just to announce the imminent completion of a **brand-new Linux release**, *which I'm calling the* **Debian Linux Release**. *[. . .]*

Debra's husband Ian A Murdock, 16/08/1993

`comp.os.linux.development`

- non-commercial distro, competitive in the OS market
- easy to install
- built collaboratively by volunteer software experts
- 1st major distro developed "openly in the spirit of GNU"

Debian: early history

1993 announcement

1994 Debian manifesto

1997 Debian Social Contract with the Free Software community

- 100% Free Software: Debian Free Software Guidelines
- give back
- don't hide problems
- priorities: users & Free Software

1998 Debian Constitution
structure and rules of a Free-Software-compatible democracy

- default: do-cracy, consensus + working code
- democracy, when needed
- scaffolding: DPL, secretary, etc.

Debian: early history of my life

1993 announcement

1994 Debian manifesto

1997 Debian Social Contract with the Free Software community
- 100% Free Software: Debian Free Software Guidelines
- give back
- don't hide problems
- priorities: users & Free Software

1998 Debian Constitution
structure and rules of a Free-Software-compatible democracy
- default: do-cracy, consensus + working code
- democracy, when needed
- scaffolding: DPL, secretary, etc.

2000 I use Debian for the first time

2004 I submit the first bug report
I contribute my first package

2006 I become an official Debian Developer

Debian, 18 years later

- $\approx$ 17'000 source packages
- $\approx$ 33'000 binary packages (amd64/sid/main)
- largest n. of ports among mainstream distros (11 official, 4 unofficial)
  - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)

Debian, 18 years later

- $\approx$ 17'000 source packages
- $\approx$ 33'000 binary packages (amd64/sid/main)
- largest n. of ports among mainstream distros (11 official, 4 unofficial)
  - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
  - The latest stable release:

FUD ALERT: 5 years ago

Debian, 18 years later

- $\approx$ 17'000 source packages
- $\approx$ 33'000 binary packages (amd64/sid/main)
- largest n. of ports among mainstream distros (11 official, 4 unofficial)
  - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
  - The latest stable release:

6.0 Squeeze, February 6th 2011

Debian, 18 years later

- $\approx$ 17'000 source packages
- $\approx$ 33'000 binary packages (amd64/sid/main)
- largest n. of ports among mainstream distros (11 official, 4 unofficial)
  - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
  - The latest stable release:

6.0 Squeeze, February 6th 2011

oldstable [still supported]: 5.0 Lenny, February 14th 2009

Debian, 18 years later

- $\approx$ 17'000 source packages
- $\approx$ 33'000 binary packages (amd64/sid/main)
- largest n. of ports among mainstream distros (11 official, 4 unofficial)
  - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
  - The latest stable release:

    6.0 Squeeze, February 6th 2011
    oldstable [still supported]: 5.0 Lenny, February 14th 2009
    old oldstable [February 15th, 2010]: 4.0 Etch, April 2nd 2007

  - Released $\approx$ 2 years
  - Security support for 3 years

Debian, 18 years later

- $\approx$ 17'000 source packages
- $\approx$ 33'000 binary packages (amd64/sid/main)
- largest n. of ports among mainstream distros (11 official, 4 unofficial)
  - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
  - The latest stable release:

    6.0 Squeeze, February 6th 2011
    oldstable [still supported]: 5.0 Lenny, February 14th 2009
    old oldstable [February 15th, 2010]: 4.0 Etch, April 2nd 2007

  - Released $\approx$ 2 years
  - Security support for 3 years
  - Upgradable

Debian, 18 years later

- $\approx$ 17'000 source packages
- $\approx$ 33'000 binary packages (amd64/sid/main)
- largest n. of ports among mainstream distros (11 official, 4 unofficial)
  - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
- 2 1/2 rolling releases: testing, unstable (sid), experimental

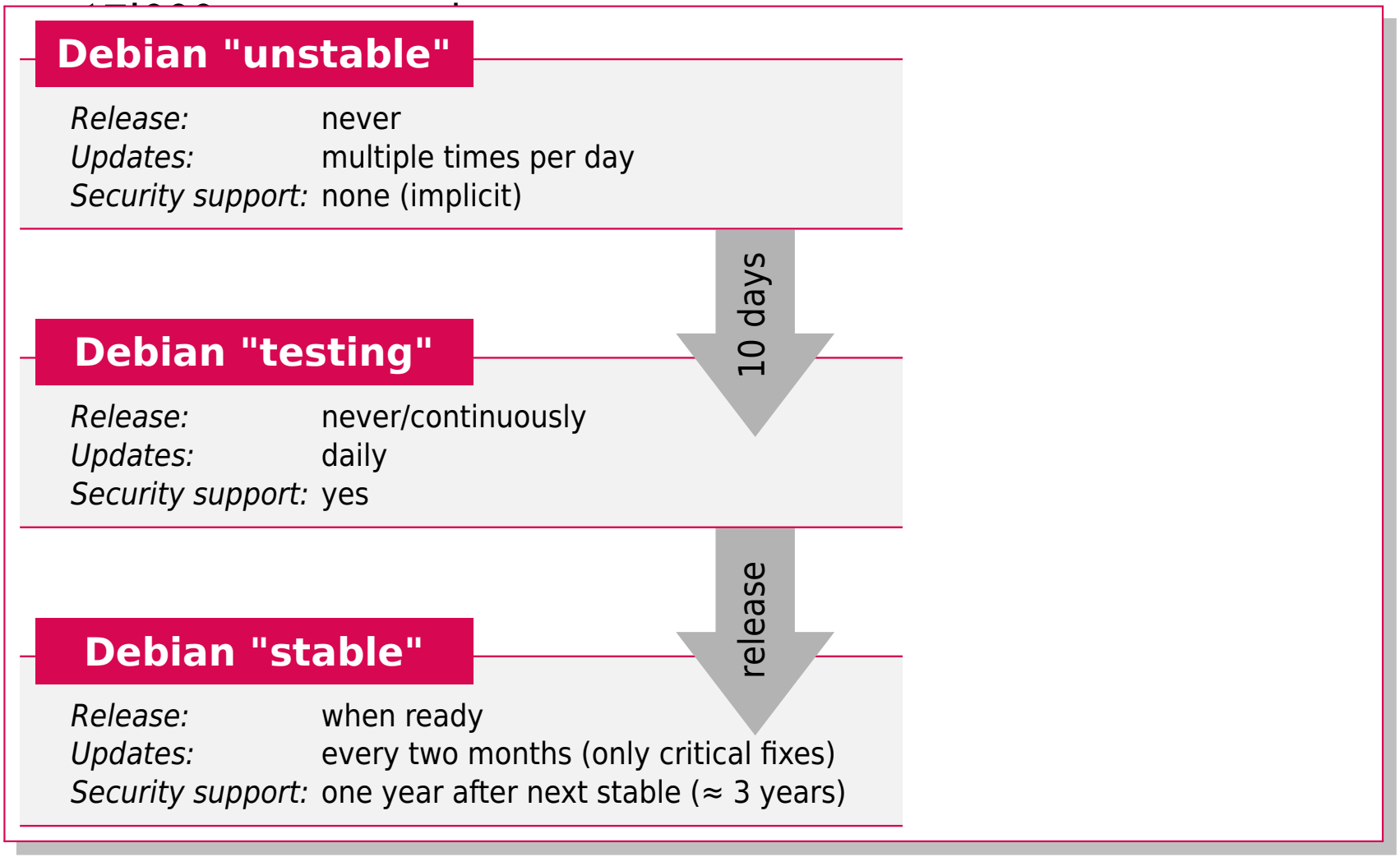
Debian, 18 years later

- [illegible]
- [illegible]
- [illegible] al)
- [illegible]
- [illegible]

Debian "stable"

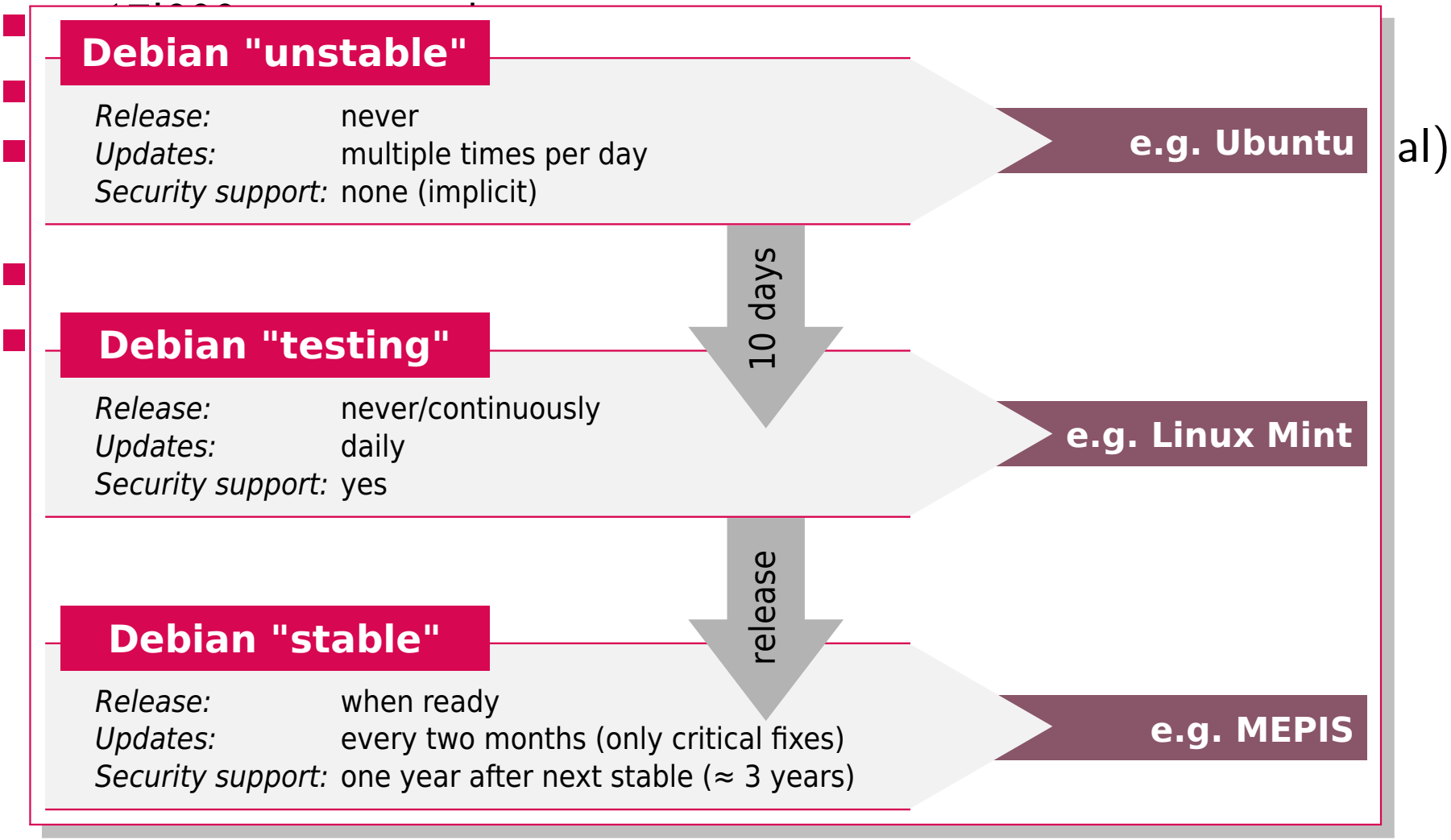
Release:	when ready
Updates:	every two months (only critical fixes)
Security support:	one year after next stable (≈ 3 years)

Debian, 18 years later



al)

Debian, 18 years later



al)

Debian, 18 years later

- ≈ 17'000 source packages
- ≈ 33'000 binary packages (amd64/sid/main)
- largest n. of ports among mainstream distros (11 official, 4 unofficial)
  - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
- 2 1/2 rolling releases: testing, unstable (sid), experimental
- 133 "derivative" distributions (source: distrowatch.com)

Debian, 18 years later

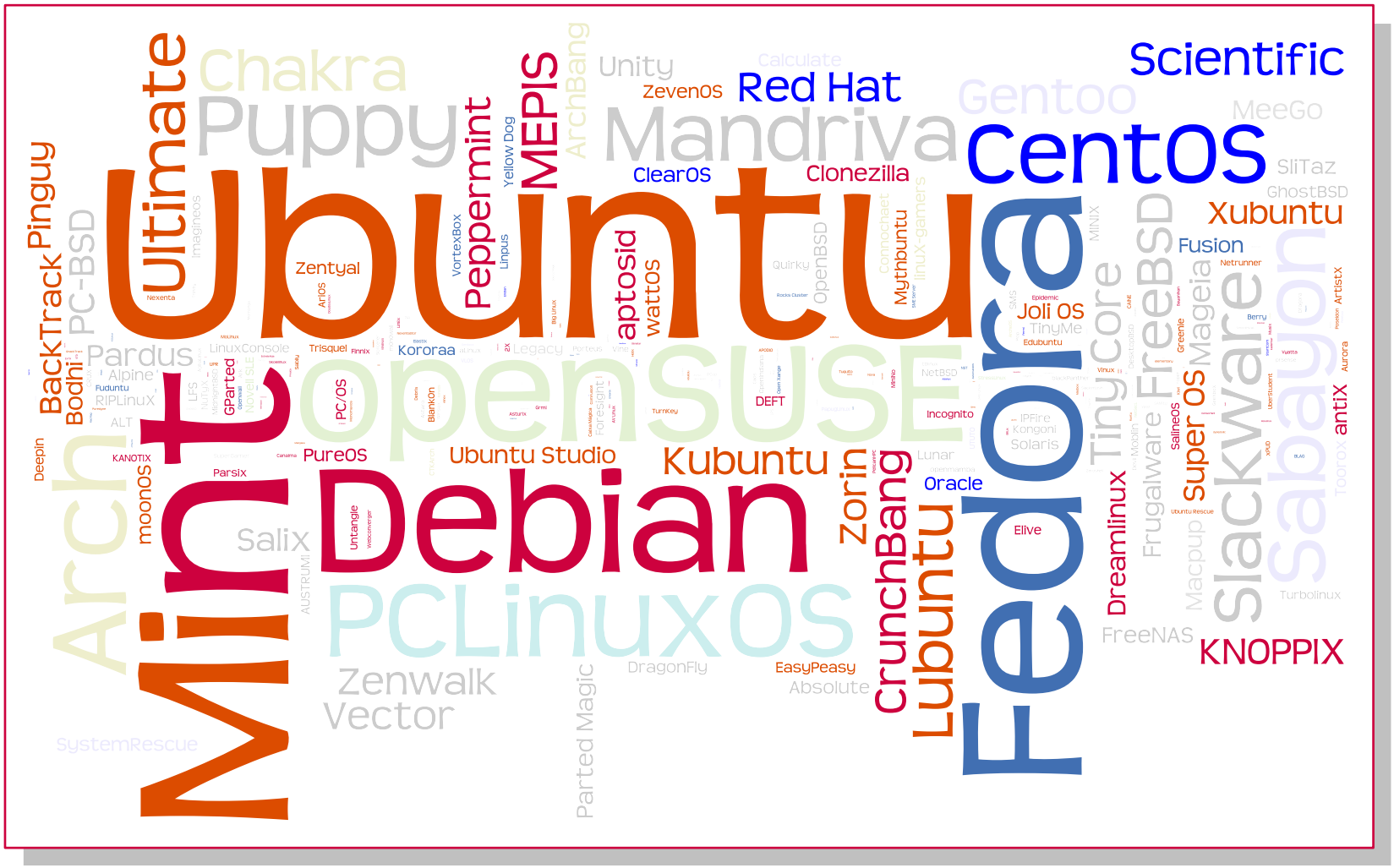
- ~ 17'000 source packages

al)

om)

Debian, 18 years later

- ~ 17'000 source packages



al)

om)

Debian, 18 years later: 133 derivatives

- ~ 17'000 source packages

al)

om)

Debian, 18 years later

- ≈ 17'000 source packages
- ≈ 33'000 binary packages (amd64/sid/main)
- largest n. of ports among mainstream distros (11 official, 4 unofficial)
  - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
- 2 1/2 rolling releases: testing, unstable (sid), experimental
- 133 "derivative" distributions (source: distrowatch.com)
- ≈ 900 active DDs + ≈200 DMs + thousands other contributors

Debian: the Universal OS

- ≈ 17'000 source packages
- ≈ 33'000 binary packages (amd64/sid/main)
- largest n. of ports among mainstream distros (11 official, 4 unofficial)
  - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
- 2 1/2 rolling releases: testing, unstable (sid), experimental
- 133 "derivative" distributions (source: distrowatch.com)
- ≈ 900 active DDs + ≈200 DMs + thousands other contributors

More on Debian & its derivatives

Who the bloody hell cares about Debian?
Stefano Zacchiroli, Debian project leader, LCA 2011, Australia

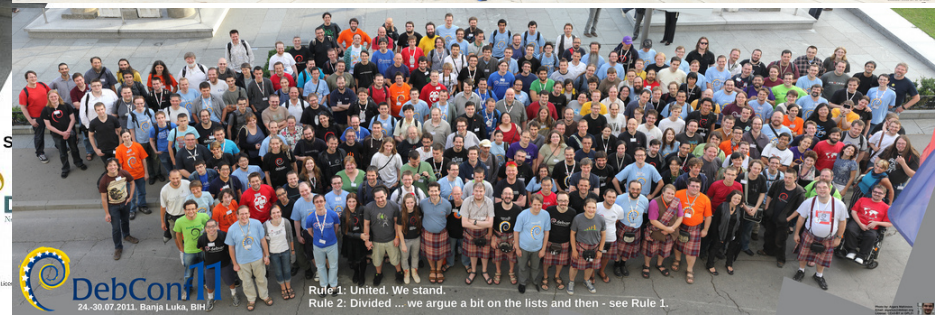
http://upsilon.cc/~zack/blog/posts/2011/01/who_the_bloody_hell_cares_about_Debian/

Debian community: DebConf 2011, 262 humanoids

≈ 900 active DDs, ≈200 DMs, >100 Teams

Debian community: DebConf 2007-2011, ≈ 900

≈ 900 active DDs, ≈200 DMs, >100 Teams



Debian: the community

≈ 900 active DDs, ≈200 DMs, >100 Teams

- Thousands of participants
- Debian developers from 57 countries
  Finland: 5,77 DDs / million
- Teams:
  `http://wiki.debian.org/Teams`
- Internationalization (i10n):
  `http://www.debian.org/international/l10n/po/rank`
- Different fields of endeavor:
  - Debian Science
  - Debian Med
  - Debichem
  - NeuroDebian
  - ...

Debian's -ocracies

1. do-ocracy

> *An individual Developer may make any technical or nontechnical decision with regard to their own work;*
>
> Debian Constitution, §3.3.1.1

2. democracy

> *Each decision in the Project is made by one or more of the following:*
>
> *1. The Developers, by way of General Resolution [...]*
>
> Debian Constitution, §2

that means:

- reputation follows work, easy to have an impact
- no benevolent dictator, no oligarchy
- no imposed decisions

  by who has money, infrastructure, employees, ...

Pythons in Debian

Python2 :

- 2.5 [in stable], 2.6 [default] and 2.7 [testing, sid]
- 2630 projects out of 16525 (16%) use Python(2)
  Used by some core tools:
  dput, git-buildpackage, piuparts, reportbug, ...
- 2.7 as the default in experimental
  http://bugs.debian.org/622279
  17 bugs holding the transition
  2.7 as the default – the release goal for wheezy

Pythons in Debian

Python2 :

- 2.5 [in stable], 2.6 [default] and 2.7 [testing, sid]
- 2630 projects out of 16525 (16%) use Python(2)
  Used by some core tools:
  dput, git-buildpackage, piuparts, reportbug, ...
- 2.7 as the default in experimental
  http://bugs.debian.org/622279
  17 bugs holding the transition
  2.7 as the default – the release goal for wheezy

Python3 :

- 3.1 [in stable], 3.2 [testing, sid]
- 65 packages dependent on Python3

Pythons in Debian

Python2 :

- 2.5 [in stable], 2.6 [default] and 2.7 [testing, sid]
- 2630 projects out of 16525 (16%) use Python(2)
  Used by some core tools:
  dput, git-buildpackage, piuparts, reportbug, ...
- 2.7 as the default in experimental
  http://bugs.debian.org/622279
  17 bugs holding the transition
  2.7 as the default – the release goal for wheezy

Python3 :

- 3.1 [in stable], 3.2 [testing, sid]
- 65 packages dependent on Python3

Maintained by :

Matthias Klose, Piotr Ożarowski, Scott Kitterman, Tollef Fog Heen, Raphael Hertzog, Michael Vogt, Marc Deslauriers, Colin Watson

Who cares about Python in Debian

111 teams :

- Debian Python Modules Team
  https://alioth.debian.org/projects/python-modules
  207 members
  maintain NumPy and SciPy
- Python Applications Packaging Team
  https://alioth.debian.org/projects/python-apps
  133 members

549 individuals

Who cares about Python in Debian

111 teams :

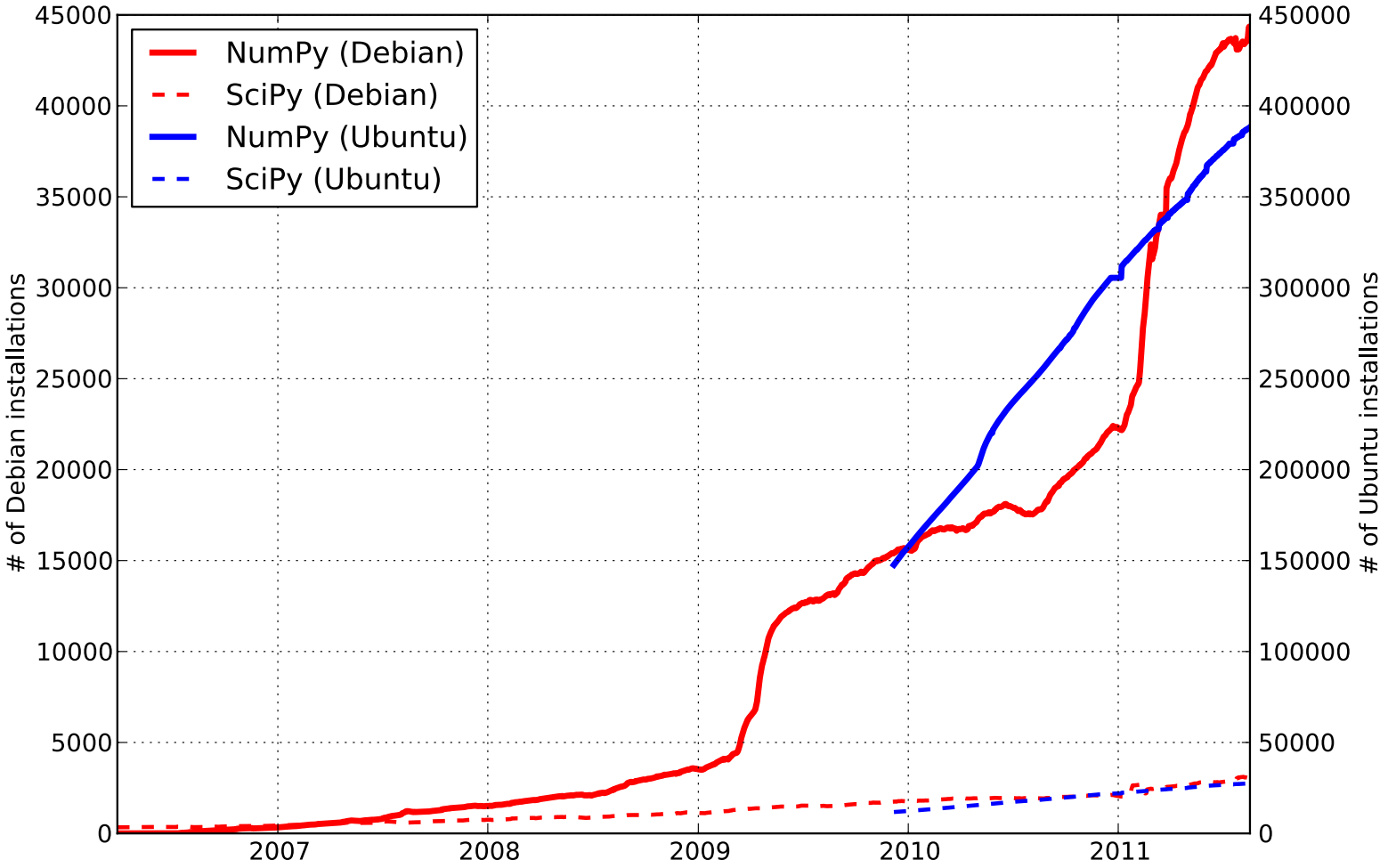
- Debian Python Modules Team
  https://alioth.debian.org/projects/python-modules
  207 members
  maintain NumPy and SciPy
- Python Applications Packaging Team
  https://alioth.debian.org/projects/python-apps
  133 members

549 individuals

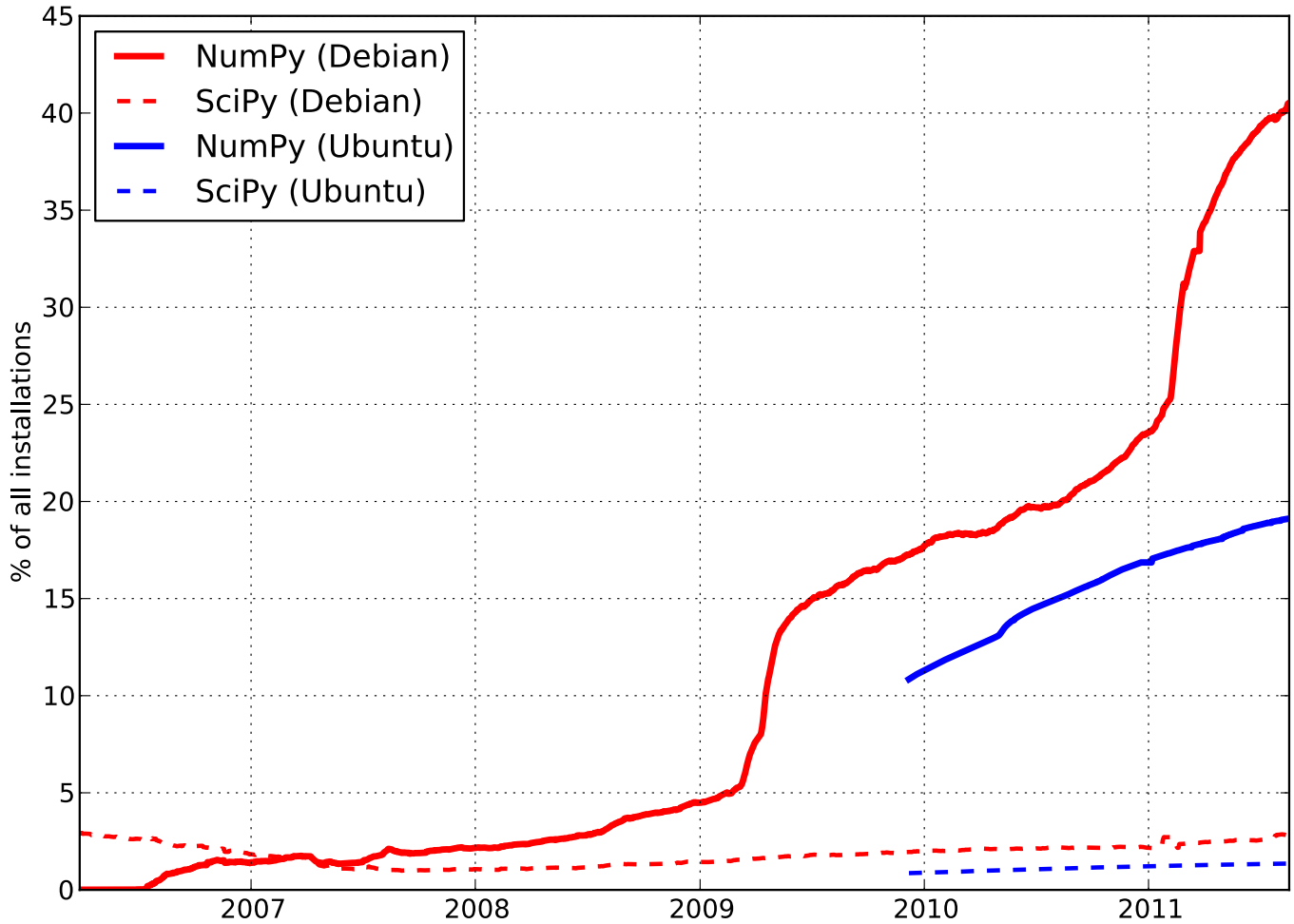
π's in Debian

- Popularity Contest
  http://popcon.debian.org

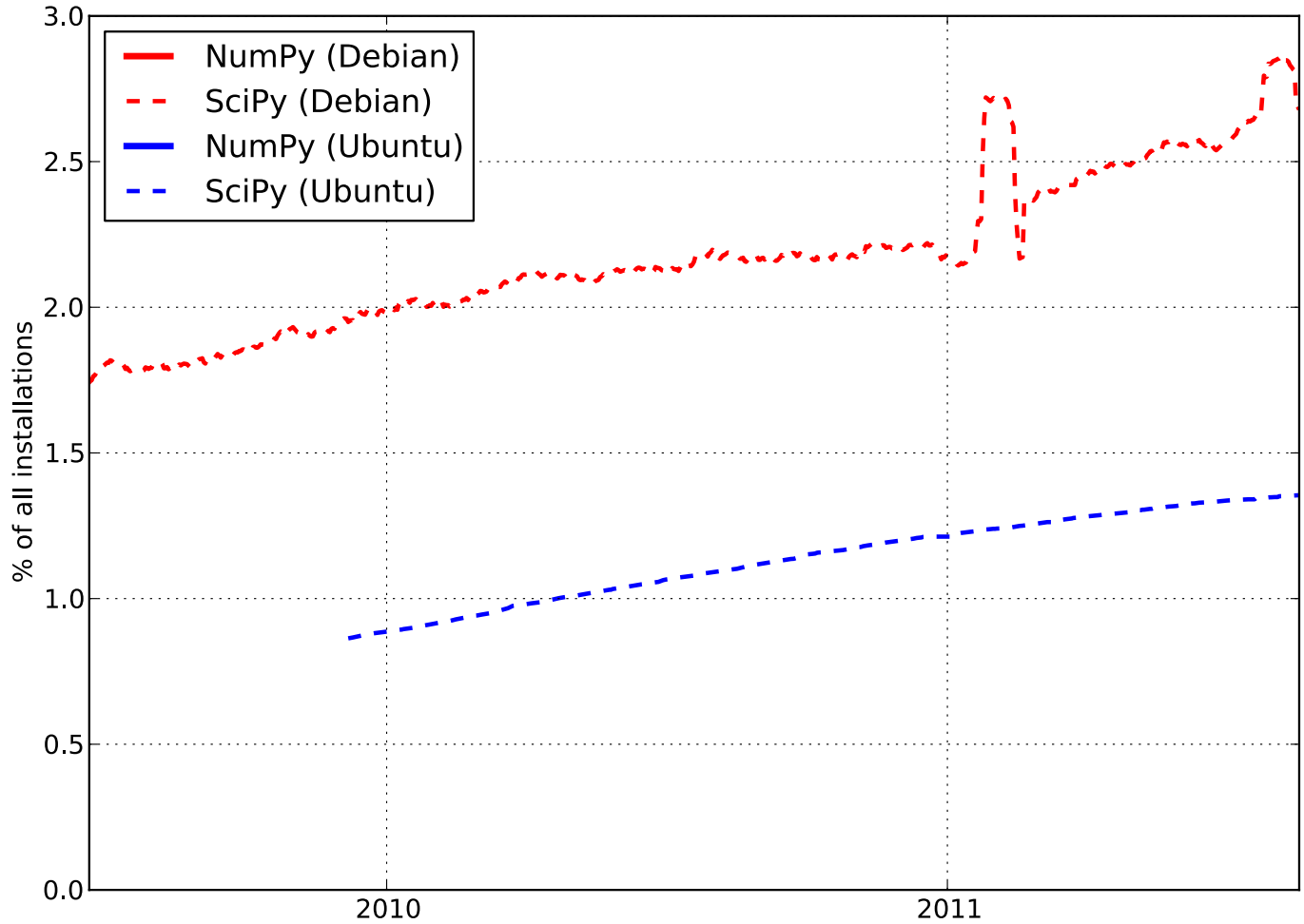
π's in Debian: popcon is growing



π's in Debian: NumPy is on 40% of Debian boxes



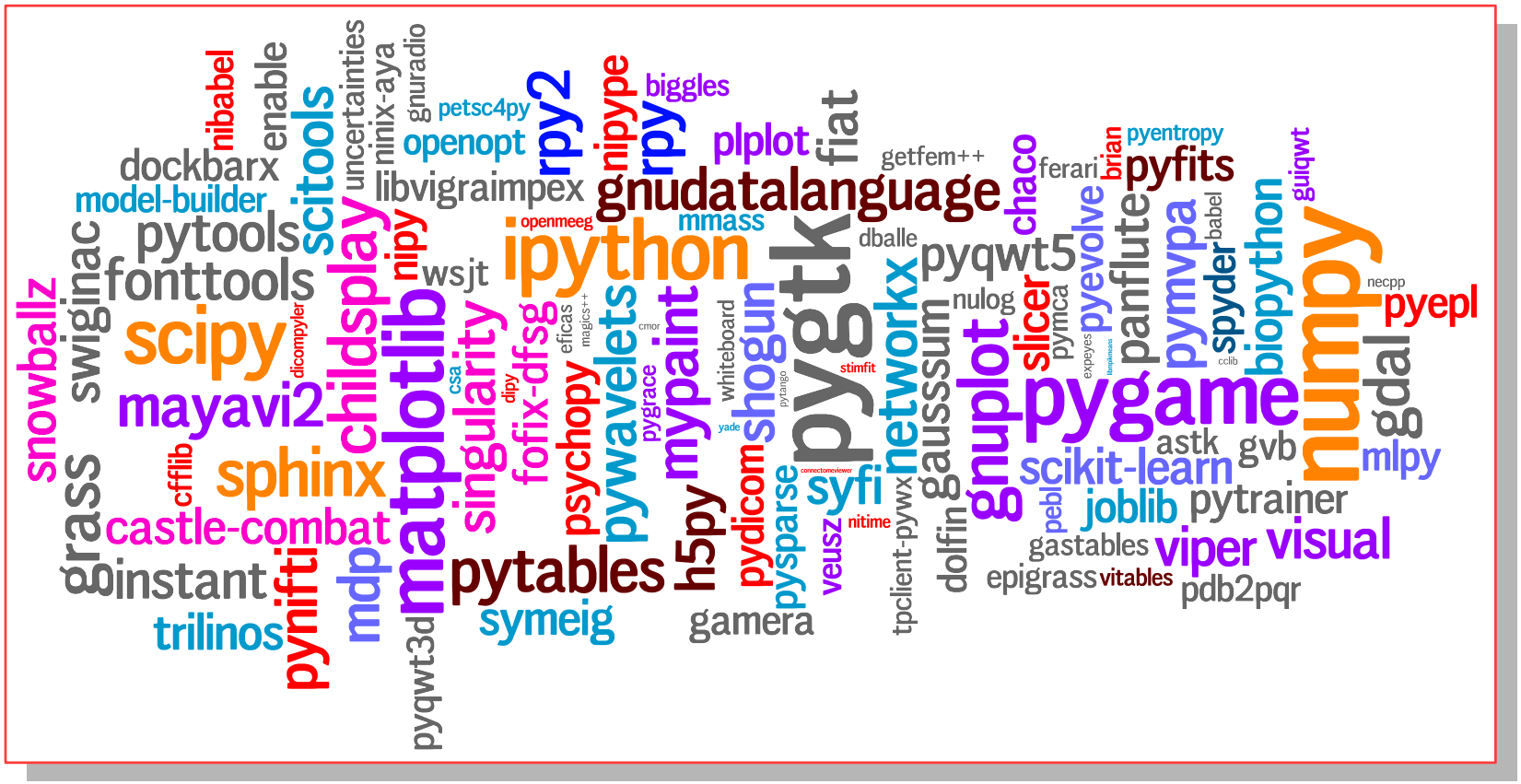
π's in Debian: SciPy is not waiting



π's in Debian

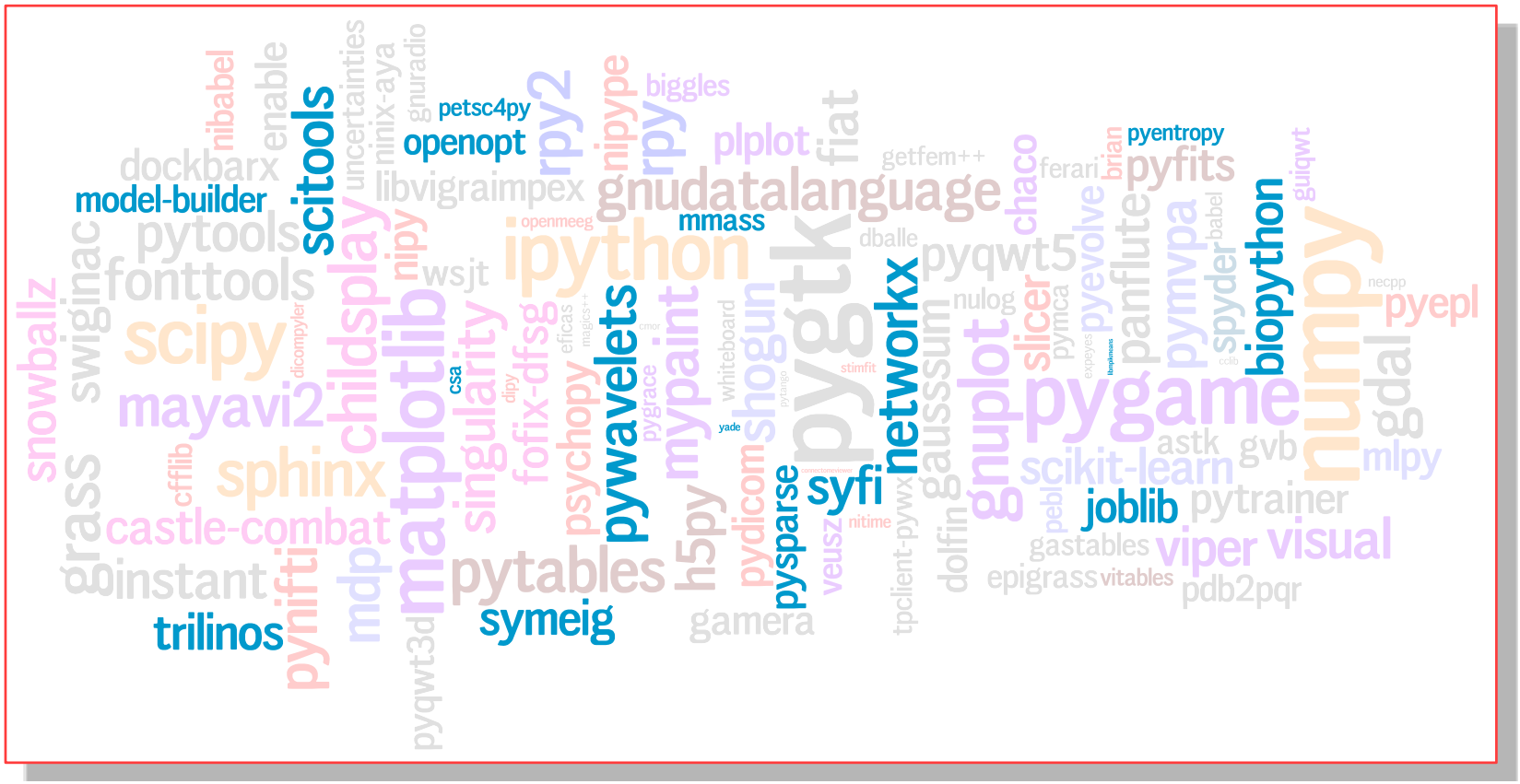
- Popularity Contest
  `http://popcon.debian.org`
  Numpy is on 40% of Debian installations
- *apt-cache rdepends python-numpy*
  - 158 packages depend on NumPy
  - 40 packages depend on SciPy

π's in Debian



π's in Debian: the Core

π's in Debian: numerical computing

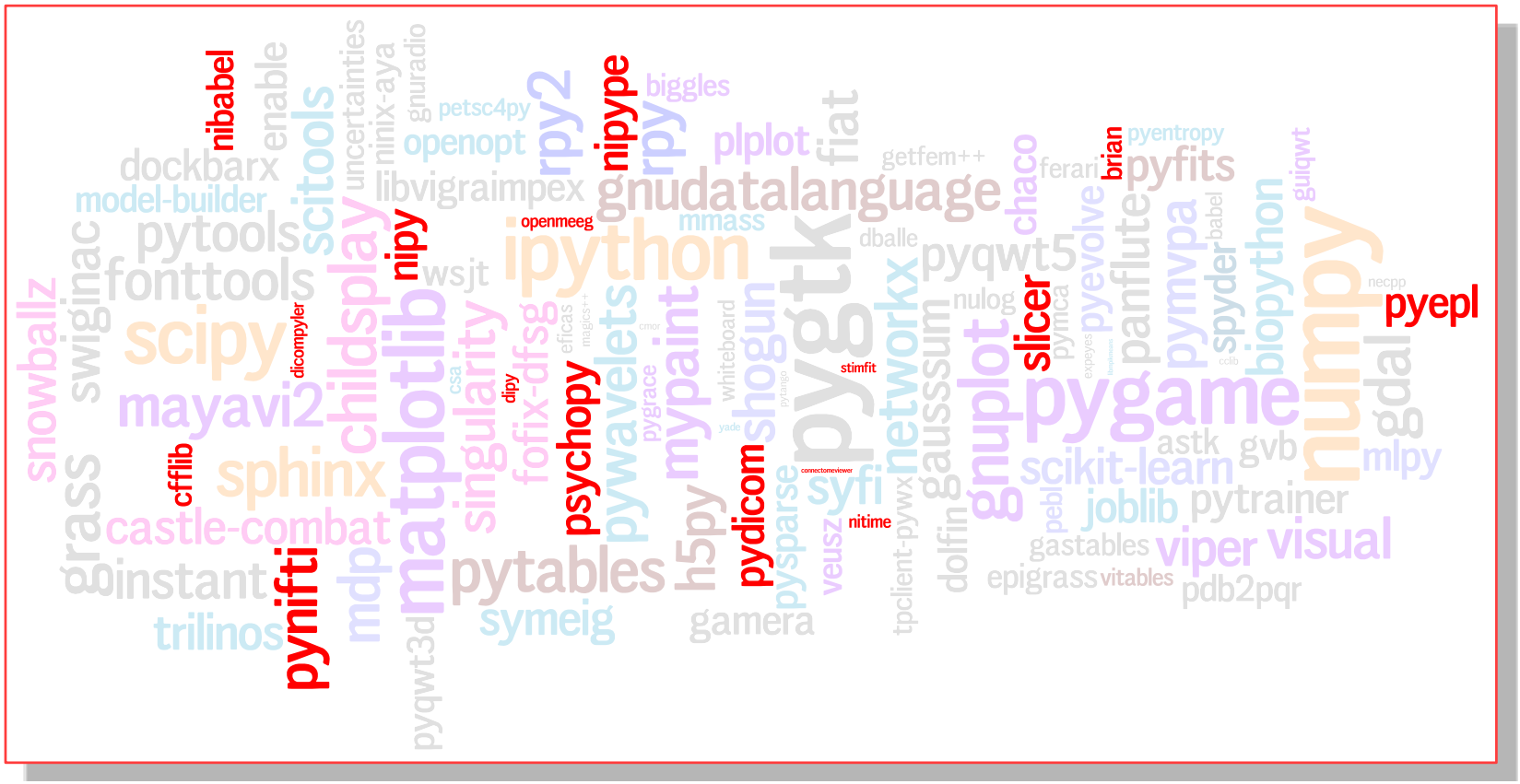


http://blends.alioth.debian.org/science/tasks/numericalcomputation

π's in Debian: machine/statistical learning

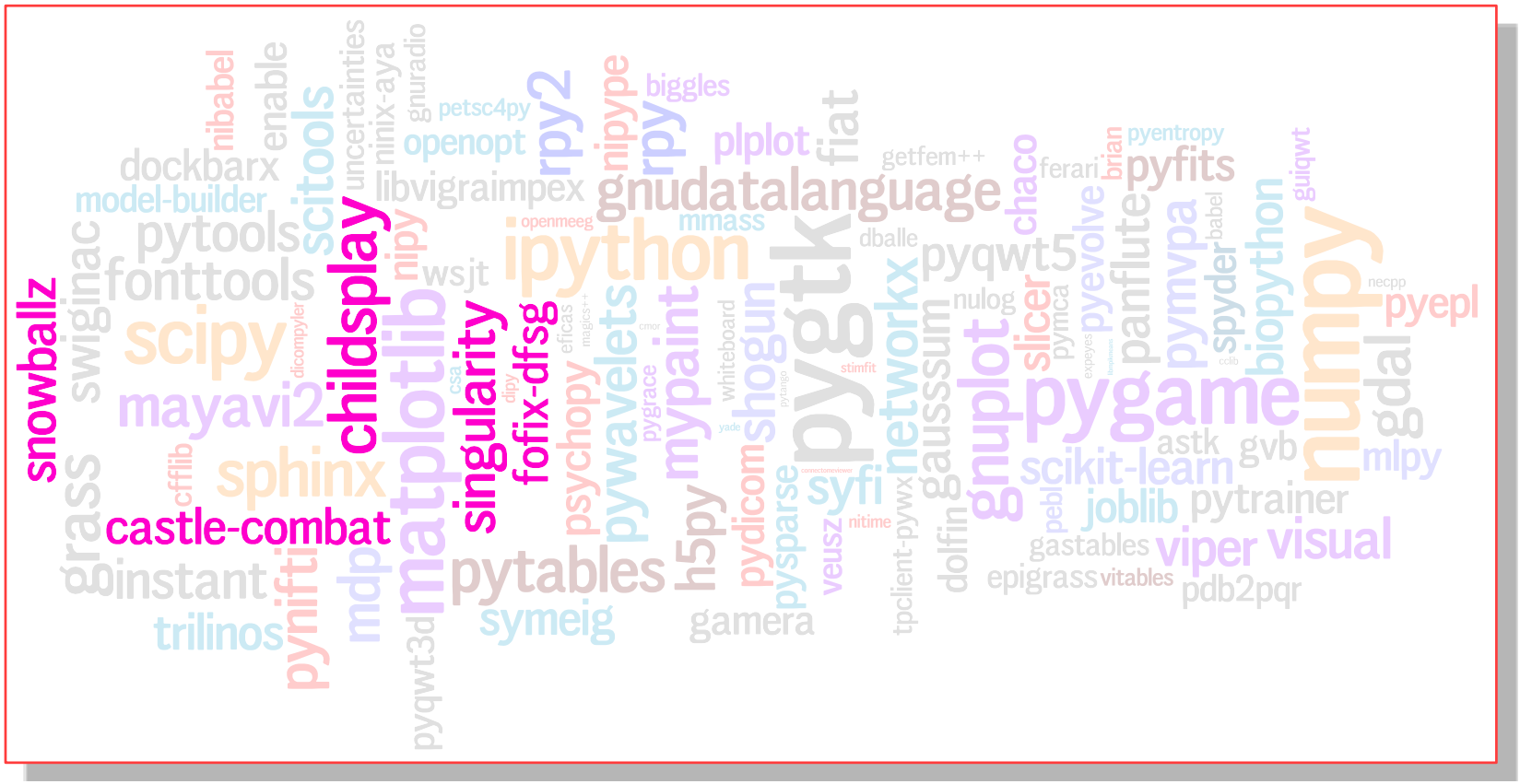
http://blends.alioth.debian.org/science/tasks/machine-learning

π's in Debian: brain sciences



http://neuro.debian.net

π's in Debian: games



π's in Debian

- Popularity Contest
  `http://popcon.debian.org`
  Numpy is on 40% of Debian installations
- *apt-cache rdepends python-numpy*
  - 158 packages depend on NumPy
  - 40 packages depend on SciPy
- Cover variety of applications out-of-the-box
- Most recent additions [Debian sid]
  - FreeNect (Free Kinect driver)
  - PyCUDA
  - PyOpenCL
  - scikits.statsmodels

Debian is beneficial for "Upstream"

Debian provides a robust deployment platform, which helps to ...

- iron out problems
  - packaged/tested on the system nearly identical to the others
  - binary builds across all supported platforms
  - (optional) package build-time (unit-)testing
  - "stable" release is stable – bugs triaged *before* the release
- deliver
  - 133 derivatives
  - Mark S.: "200 millions of Ubuntu users in 3 years and 9 months"
  - official Debian mirrors in 46 countries
    http://www.debian.org/mirror/list
- engage more caring hands
  - QA activities: archive rebuilds (FTBFS), package QA tools
  - centralized transitions
- report usage statistics
  http://popcon.debian.org

Debian is beneficial for "Upstream"

Debian provides a robust deployment platform, which helps to ...

- iron out problems
  - packaged/tested on the system nearly identical to the others
  - binary builds across all supported platforms
  - (optional) package build-time (unit-)testing
  - stable release is stable – bugs triaged *before* the release
- deliver
  - 133 derivatives
  - Mark S.: "200 millions of Ubuntu users in 3 years and 9 months"
  - official Debian mirrors in 46 countries
    `http://www.debian.org/mirror/list`
- engage more caring hands
  - QA activities: archive rebuilds (FTBFS), package QA tools
  - centralized transitions
- report usage statistics
  `http://popcon.debian.org`

> I have always found my friends Debian developers to be pretty good at getting me do boring but useful stuff.
>
> –Gael Varoquaux

Help yourself to help Debian

- Have a deterministic version
- Be conscious about *all* licenses
- Allow for modularity
  - use (documented) "standard" build mechanisms
  - treat 3rd party as 3rd party
    - no forks – forward fixes upstream, request bugfix releases
    - allow an option to build against system-installed versions
  - be compatible with recent released versions NumPy/SciPy
- Be prepared for feedback
- Provide unit-/doc- tests and examples
  - easy way to run only lightweight portion (for build-time testing)
  - conventional means to run the tests
  - use tempfile.* instead of the work-tree
  - do not hardcode matplotlib backends (unless required)
- Test/use your software on Debian
  (especially during Debian freeze)

---

http://neuro.debian.net/blog/2011/2011-08-23_getting_stuff_packaged.html

Debian is a rich platform for Python development

Debian facilitates software development by providing out-of-the-box...

- Multiple supported versions of Python
- Python Editors/IDEs/refactoring tools
  - vim, emacs (GNU Python mode, python-mode, ropemacs)
  - DrPython, Eric, Geany, gEcrit, PIDA, Spyder, ...
  - pylint, pyflakes
  - rope, bicyclerepair
- debugging facilities
  - pdb, pydb, pudb, winpdb
  - advanced extensions debugging (in a minute)
- easy ways to bootstrap a complete system (in 2 minutes)

Advanced extensions debugging facilities

GDB	inspect Python stack
Valgrind	pin-point segfaults and memory leaks
Profiling GUI	kcachegrind, hotshot + kcachegrind-converters
	https://github.com/PyMVPA/PyMVPA/blob/master/tools/profile
DMTCP	snapshot lengthy computations
FReD	[coming] reversible debugger

GDB: inspect Python stack

```
> gdb --args /usr/bin/python-dbg segfault.py
GNU gdb (GDB) 7.3.50.20110627-cvs-debian
...
Program received signal SIGSEGV, Segmentation fault.
...
(gdb) py <TAB>
py-bt       py-down     py-list     py-locals   py-print    py-up
```

GDB: Python stack

```
(gdb) bt
...
#10  ... /arrayprint.py, line 156, in _leading_trailing ...
#11  ... at ../Python/ceval.c:3836 ...
(gdb) py-bt
#10 Frame 0xf5c6d0, ... /arrayprint.py, line 156 ....
                                        a[-_summaryEdgeItems:]))
#13 Frame 0xf63230, ... ./arrayprint.py, line 162 ...
                min(len(a), _summaryEdgeItems))]
(gdb) py-up
#13 Frame 0xf63230 ...
                min(len(a), _summaryEdgeItems))]
(gdb) py-down
#10 Frame 0xf5c6d0 ...
                            a[-_summaryEdgeItems:]))
```

DMTCP: Snapshot your Python

Why?

- Stop/resume the lengthy task
  - across power outages
  - "reversible" debugging
- Move the task across identical boxes (PBS, Condor)

- Distributed MultiThreaded CheckPointing
- Works with Linux kernel 2.6.9 and later
- Supports sequential and multi-threaded computations across single/multiple hosts
- Entirely in user space (no kernel modules or root privilege)
- Transparent (no recompiling, no re-linking)
- DMTCP Team centered around Northeastern U., with collaborators from MIT and Siberian State U. of Telecom. and Informatics
- Available in Debian $\geq$ wheezy (current testing)

DMTCP: Snapshot your Python

Why?

- Stop/resume the lengthy task
  - across power outages
  - "reversible" debugging
- Move the task across identical boxes (PBS, Condor)

STANDALONE USAGE

```
dmtcp_checkpoint a.out

dmtcp_command --checkpoint

dmtcp_restart ckpt_a.out_*.dmtcp
```

DMTCP: Snapshot your Python

Why?

- Stop/resume the lengthy task
  - across power outages
  - "reversible" debugging
- Move the task across identical boxes (PBS, Condor)

Python interface

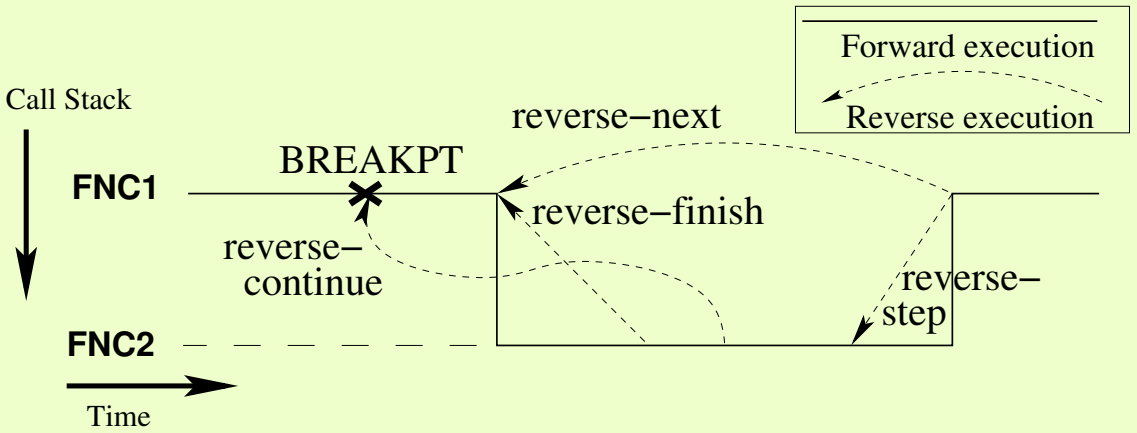
In the next release

DMTCP: Snapshot your Python

Why?

- Stop/resume the lengthy task
  - across power outages
  - "reversible" debugging
- Move the task across identical boxes (PBS, Condor)

FReD: Fast Reversible Debugger (WiP)



http://www.cs.wisc.edu/condor/CondorWeek2011/wednesday_condor.html

Bootstrapping a complete Debian system

Why?

- build/test/use previous or upcoming Debian (or Ubuntu) release
- get clean environment (track dependencies)
- mimic user's setup

Bootstrapping a complete Debian system

Why?

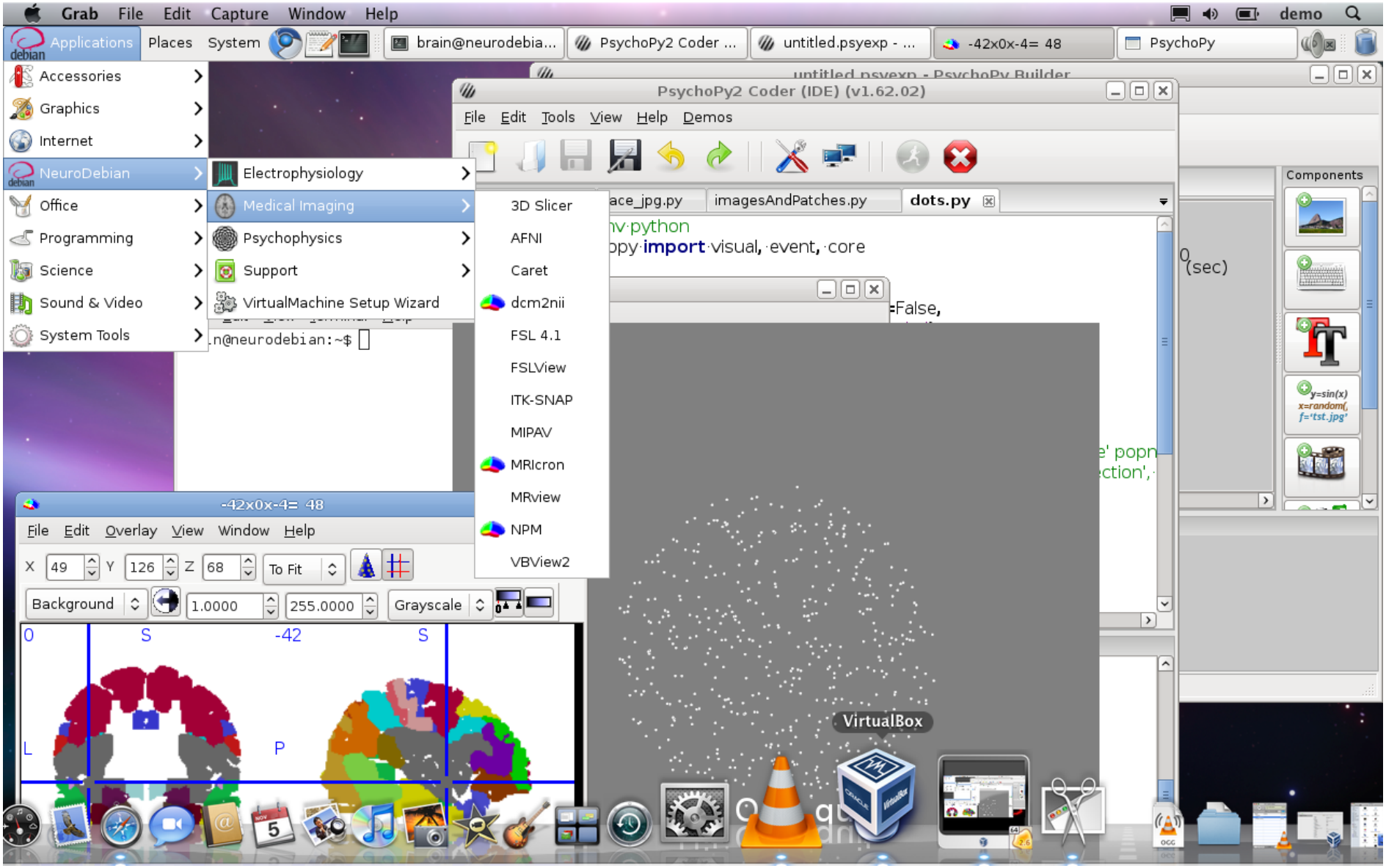
- build/test/use previous or upcoming Debian (or Ubuntu) release
- get clean environment (track dependencies)
- mimic user's setup

- debootstrap + schroot: install into any directory
- vmdebootstrap: generate a virtual machine
  http://blog.liw.fi/posts/vmdebootstrap/
- Fully Automated Installation (FAI): http://fai-project.org/
- VirtualBox: install or use pre-crafted virtual appliance
- Cloud: http://wiki.debian.org/Cloud

debootstrap + schroot

```
> debootstrap sid /var/cache/chroots/sid-amd64
> sudo bash -c "cat << EOF >| /etc/schroot/chroot.d/sid-amd64
[sid-amd64]
description=Debian sid (forever unstable) [amd64]
type=directory
location=/var/cache/chroots/sid-amd64
users=YOURLOGIN
aliases=unstable,sid,default
EOF"
> schroot
```

NeuroDebian VM in VirtualBox



http://neuro.debian.net/vm.html

Ways to contribute

http://wiki.debian.org/ProjectNews/HowToContribute
http://raphaelhertzog.com/2011/06/30

- reportbug (+ patches)
- Internationalization (i18n):
  http://www.debian.org/doc/manuals/intro-i18n
- packaging
  - Luca's tutorial
    apt-get install packaging-tutorial
    http://www.lucas-nussbaum.net/blog/?p=640
  - Bootstrap packaging of Python modules:
    py2dsc (python-stdeb package)
  - Good night reading: Debian Policy
  - Seek mentor/sponsor-ship: http://mentor.debian.org
  - Become "Debian Maintainer":
    http://wiki.debian.org/DebianMaintainer
  - Become "Debian Developer":
    http://wiki.debian.org/DebianDeveloper


Brain Download:
iz compltes.
ICANHASCHEEZBURGER.COM

Acknowledgements

Michael Hanke

Free and opensource software developers

Debian Community

Python/NumPy/Scipy Maintainers

Jim Haxby

Thanks!

Yaroslav O. Halchenko

yoh@debian.org

http://www.onerussian.com

How many care about Python

```
> grep-dctrl -s Maintainer -F Build-Depends python
    -o -F Build-Depends python-dev
-o -F Build-Depends python-all
/var/lib/apt/lists/*\_sid\_main\_source\_Sources
  | sort | uniq -c | sort -n -r
  | wc -l
660
```

Who cares: teams

```
> grep-dctrl -s Maintainer -F Build-Depends python
    -o -F Build-Depends python-dev
-o -F Build-Depends python-all
/var/lib/apt/lists/*\_sid\_main\_source\_Sources
  | sort | uniq -c | sort -n -r
  | grep -e alioth -e team -e Maintainers -e Debian
  | wc -l
111
```

Who cares: teams

```
> ...
  | grep -e alioth -e team -e Maintainers -e Debian | head
    232 Maintainer: Debian Python Modules Team <python-mod...
     57 Maintainer: Debian Tryton Maintainers <tryton@list...
     51 Maintainer: Debian OLPC <debian-olpc-devel@lists.a...
     45 Maintainer: Python Applications Packaging Team <py...
     40 Maintainer: Debian/Ubuntu Zope Team <pkg-zope-deve...
     40 Maintainer: Debian Multimedia Maintainers <pkg-mul...
     26 Maintainer: NeuroDebian Team <team@neuro.debian.ne...
     26 Maintainer: Debian Science Maintainers <debian-sci...
     26 Maintainer: Debian QA Group <packages@qa.debian.or...
     26 Maintainer: Debian Bazaar Maintainers <pkg-bazaar-...
```

Who cares: individuals

```
> grep-dctrl -s Maintainer -F Build-Depends python
    -o -F Build-Depends python-dev
-o -F Build-Depends python-all
/var/lib/apt/lists/*\_sid\_main\_source\_Sources
  | sort | uniq -c | sort -n -r
  | grep -v -e alioth -e team -e Maintainers -e Debian
  | wc -l
549
```

Who cares: individuals

```
> ...
  | head
26 Maintainer: Matthias Klose <doko@debian.org>
16 Maintainer: David Paleino <dapal@debian.org>
14 Maintainer: Arnaud Fontaine <arnau@debian.org>
13 Maintainer: Jelmer Vernooij <jelmer@debian.org>
12 Maintainer: Pierre Chifflier <pollux@debian.org>
12 Maintainer: Josselin Mouette <joss@debian.org>
11 Maintainer: Georges Khaznadar <georgesk@ofset.org>
11 Maintainer: Chris Lamb <lamby@debian.org>
11 Maintainer: Alessio Treglia <alessio@debian.org>
10 Maintainer: Scott Kitterman <scott@kitterman.com>
```

NumPy: who cared

```
/usr/share/doc/python-numpy/changelog.Debian.gz
     22 Marco Presi (Zufus)
     20 Ondrej Certik
     13 Kumar Appaiah
      9 Sandro Tosi
      9 Matthias Klose
      2 Tiziano Zito
      ... 15 more ...
```

SciPy: who cared

```
/usr/share/doc/python-scipy/changelog.Debian.gz
     15 Marco Presi (Zufus)
     14 Ondrej Certik
      7 José Fonseca
      7 Alexandre Fayolle
      5 Luca Falavigna
      4 Sandro Tosi
      3 Varun Hiremath
      3 Piotr Ożarowski
      2 Matthias Klose
      ... 5 more ...
```