Open is not enough: benefits from Debian as an integrated, community-driven computing platform

Yaroslav O. Halchenko & Michael Hanke

Debian Project

Dartmouth College, New Hampshire
University of Magdeburg, Germany

SEA 2013, Boulder CO

March 2 2013

Am I the right person?

To: debian-user@lists.debian.org

Subject: Duplicating Debian Installations? From: tres@rap.ucar.EDU (Tres Hofmeister) Date: Mon, 19 Aug 1996 19:10:41 -0600 (MDT)

Can someone point me towards info. on how one can duplicate the packages installed on one Debian system on another? Going through dselect by hand seems just a bit too tedious for multiple installations... Thanks.

Tres Hofmeister tres@ncar.ucar.edu

Research Applications Program National Center for Atmospheric Research

https://lists.debian.org/debian-user/1996/08/msg00890.html

Based on a true story...



4.511 Views



I and



http://www.pymvpa.org

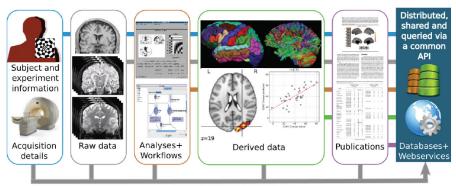


http://www.debian.org



http://neuro.debian.net

Our domain



Provenance

Poline et. al, Data sharing in neuroimaging research, 10.3389/fninf.2012.00009

Our belief #1

There is **no** alternative to free and open-source software for scientific research.

Common problem: Availability != Accessibility

Common problem: Availability != Accessibility

"Standard" model of scientific software deployment

To download and install A on your system, you need to:

- Register/Request an account from the B
- Read these notes for C and D systems
- Download the source code and/or appropriate binaries for your system from E
- Remove old versions
- Install software
- Test software

Common problem: Availability != Accessibility

"Standard" model of scientific software deployment

To download and install A on your system, you need to:

- Register/Request an account from the B
- Read these notes for C and D systems
- Download the source code and/or appropriate binaries for your system from E
- Remove old versions
- Install software
- Test software
- If you decide to go forth with building from source code, good luck!

Possible user experiences

Not so positive

- Software developers have no access to my system X and I need to build a few (dozens) of additional pre-requisites first:
 - I better ask our IT personnel to deploy it usually takes less than a month
 - PhD students have all the time in the universe.

Possible user experiences

Not so positive

- Software developers have no access to my system X and I need to build a few (dozens) of additional pre-requisites first:
 - I better ask our IT personnel to deploy it usually takes less than a month
 - PhD students have all the time in the universe . . .

Positive

Only an hour and "A" seems to be running – I can get back to do research again!

Possible user experiences

Not so positive

- Software developers have no access to my system X and I need to build a few (dozens) of additional pre-requisites first:
 - I better ask our IT personnel to deploy it usually takes less than a month
 - PhD students have all the time in the universe . . .

Positive

- Only an hour and "A" seems to be running I can get back to do research again!
- Student's life is short:
 - Having A on one box should be enough
 - I better not have to upgrade it ever again
 - I will not even think about F and G alternatives
 - I value others' time: let me blog a "HOWTO install A v. x.y.z on C"

Hanke, M. (2012). Share your tools! But fear the wombat! http://youtu.be/8t6znEOEDVo

Not so positive

- Software developers have no access to my system X are need to build a few (dozens) of additional pre-requisite first:
 - I better ask our IT personnel to deploy it itakes it whan a month
 - PhD students have all the time the inverse...

Positive

- Only an hour and seein, to be runesearch again.
- State life is 17th
 - A on one box should be end
 - ternot have to upgrade it ever a
 - I will not even think about F and G alt
 - I value others' time: let me blog a "HO stall A C"

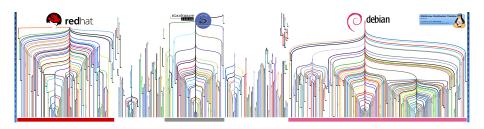
Scientists report:

GNU/

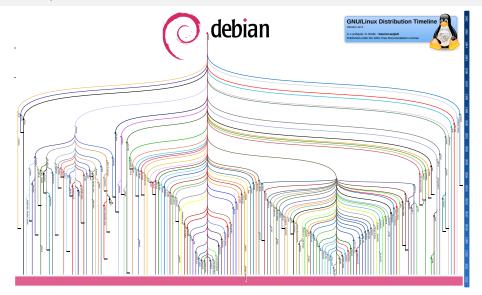
Linux users report the lowest average time they need to invest in maintenance of their personal computing environment (5.77 h/month).

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

GNU/Linux'es



GNU/Linux'es



http://futurist.se/gldt

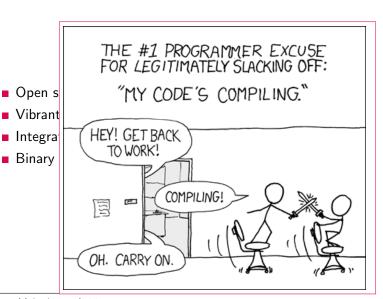
Our belief #2

Debian provides the best proven, scalable, and sustainable approach toward free and open-source software platform for scientific research

Debian's recipe

- Open standards
- Vibrant community: democracy and do-o-cracy
- Integration starting with the core OS
- Binary distribution based on source packages

Debian's recipe: Binary distribution



http://xkcd.com/303

Debian's recipe

- Open standards
- Vibrant community: democracy and do-o-cracy
- Integration starting with the core OS
- Binary distribution based on source packages
- Extensible distribution and archival infrastructure

Legal assurance Was: Register/Request an account from the B

3 Debian archive areas

- main software under DFSG compliant licenses ¹
- contrib DFSG-wannabees (depend on non-free) or helpers (e.g., matlab-support)
- non-free (re-)distributable but restricted

¹See Debian Social Contract http://www.debian.org/social_contract

Legal assurance Was: Register/Request an account from the B

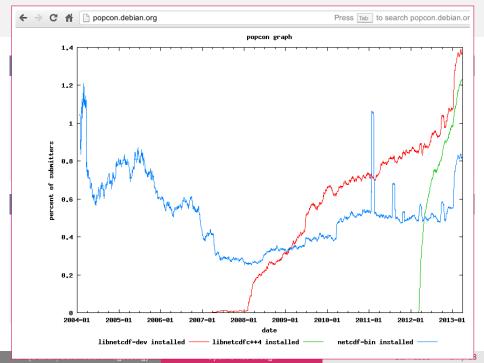
3 Debian archive areas

- main software under DFSG compliant licenses ¹
- contrib DFSG-wannabees (depend on non-free) or helpers (e.g., matlab-support)
- non-free (re-)distributable but restricted

Pros

- you know that everything is free and open-source
- eases integration
- helps to assure longevity
- http://popcon.debian.org is there to replace "registration" pages

¹See Debian Social Contract http://www.debian.org/social_contract



Legal assurance Was: Register/Request an account from the B

Recommendations

- use standard OSI (http://opensource.org) and DFSG compliant licenses
- do not restrict the domain of application without need
- do not insert code/data with questionable license terms
- maintain a top-level list of licenses used in you project
- beware-of and/or state your trademark policies

Building

Building: developers



Image by Lewis Hine, 1930

Building: well-engineered software



Image by Lewis Hine, 1930 Image by Ad Meskens

Building : "state-of-art"



- Debian packaging strictly follows Debian Policy¹
- 3rd-party modules become 1st-class citizens
- Figure out dependencies once specify in debian/control
- Figure out building/installation once specify in debian/rules

¹http://www.debian.org/doc/debian-policy/

- Debian packaging strictly follows Debian Policy¹
- 3rd-party modules become 1st-class citizens
- Figure out dependencies once specify in debian/control
- Figure out building/installation once specify in debian/rules

Pros

- anyone can build any PACKAGE: apt-get source -b PACKAGE
- simplifies contributions
- allow for archive-wide QA rebuilds
 - library migrations
 - new compilers
 - new ports

¹http://www.debian.org/doc/debian-policy/

Recommendations

- keep source distribution modular:
 - code-vs-data
 - keep 3rd party as 3rd party do not clone
- have a deterministic version
- provide exhaustive specification of (build-)dependencies
- assure API/ABI compatibility if you deliver libraries:
 - http://upstream-tracker.org/
 - http://github.com/lvc/abi-compliance-checker

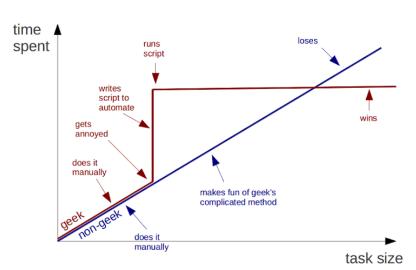
Recommendations

- keep source distribution modular:
 - code-vs-data
 - keep 3rd party as 3rd party do not clone
- have a deterministic version
- provide exhaustive specification of (build-)dependencies
- assure API/ABI compatibility if you deliver libraries:
 - http://upstream-tracker.org/
 - http://github.com/lvc/abi-compliance-checker

Above recommendations will help to deliver your software to any integration/distribution platform, not just Debian.

Testing

Geeks and repetitive tasks



debian/rules: can exercise tests while building the binary packages

debian/rules: can exercise tests while building the binary packages

Pros

- obtain a CI farm with > 13 architectures/ports
- have I mentioned archive wide QA rebuilds?
 Go ahead of the OS not behind
- preclude problems on users' deployments

debian/rules: can exercise tests while building the binary packages

Pros

- obtain a CI farm with > 13 architectures/ports
- have I mentioned archive wide QA rebuilds?
 Go ahead of the OS not behind
- preclude problems on users' deployments
- autopkgtest: automatic as-installed package testing

Pros

- exercise testing on installed machines
- could be "heavy" tests

Recommendations

- share your unit-, regression-, integration- tests
- work on your tests coverage

Recommendations

- share your unit-, regression-, integration- tests
- work on your tests coverage
- more problems we catch before delivering to users less of WOMBAT

Install simple editor

apt-get install gedit

Install simple editor

apt-get install gedit

Install workload management system

apt-get install condor

Install simple editor

apt-get install gedit

Install workload management system

apt-get install condor

Install the entire suite of packages for meteorology

apt-get install science-meteorology

Install simple editor

apt-get install gedit

Install workload management system

apt-get install condor

Install the entire suite of packages for meteorology

apt-get install science-meteorology

Keep the whole system up-to-date

apt-get update && apt-get upgrade

Debian Science: http://wiki.debian.org/DebianScience http://blends.alioth.debian.org/science/tasks/meteorology

Task pages: astronomy, astronomy-dev, bci, biology, chemistry, dataacquisition, dataacquisition-dev, distributed computing, electronics, electrophysiology, engineering, engineering-dev, geography, highenergy-physics, highenergy-physics-dev, imageanalysis, linguistics, machine-learning, mathematics, mathematics-dev, meteorology, meteorology-dev, nanoscale-physics, nanoscale-physics-dev, neuroscience-cognitive, neuroscience-datasets, neuroscience-modeling, numerical computation, physics, physics-dev, psychophysics, robotics, simulations, statistics, typesetting, viewing





Meteorology

Debian Science Meteorology packages

This metapackage is part of the Debian Pure Blend "Debian Science" and installs packages related to Meteorology and Climate. The list to the right includes various

software projects which are of some interest to the Debian Science Project. Currently, only a few of them are available as Debian packages. It is our goal, however, to include all software in Debian Science which can sensibly add to a high quality Dehian Pure Blend For a better overview of the project's availability as a Debian package. each head row has a color code according to this scheme:

- Official Debian packages with high relevance
- Official Debian packages with lower relevance
- No known packages available but some record of interest (WNPP bug)

If you discover a project which looks like a good candidate for Debian Science to you, or if you have prepared an unofficial Debian package, please do not hesitate to send a description of that project to the Debian Science mailing list

Debian Science Meteorology packages

Official Debian packages with high relevance **Aweather** Advanced Weather Monitoring Popcon: 16 users (5 upd.) Program License: DESG free http://lug.rose-Versions and Archs Official Debian package hulman.edu/proj/aweather Edit Debtags Maintainer: Debian Science Maintainers (Andy Spencer) AWeather is an advanced weather program which is designed to be used by weather enthusiasts. AWeather is not a weather dockapp that simply displays a pre-computed forecast. It is designed to be an easy-Upload screenshot to-use program that integrates a variety of weather data in a simple unified interface. AWeather currently supports radar and weather alerts from the United State National Weather Service Popcon: 31 users (3 upd.) Cdo Climate Data Operators License: DFSG free Versions and Archs Official Debian package https://code.zmaw.de/projects/cdo Maintainer: Alastair McKinstry Edit Debtags Climate Data Operators are a collection of command line Operators to manipulate and analyse Climate model Data. Supported data formats Upload screenshot are GRIB, netCDF, SERVICE, EXTRA and IEG. There are more than 400 operators available.

Cmip5-cmor-tables

Maintainer: Alastair McKinstry

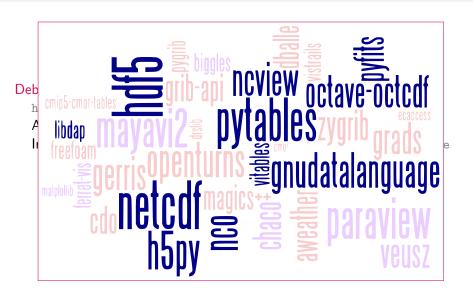
Popcon: 0 users (0 upd.)* CMIP5 tables for the Climate Model Output Rewriter library http://www2-pcmdi.llnl.gov/cmor

Versions and Archs Go tagging

License: DESG free Official Debian package Git

SEA 2013









```
Debian Science: http://wiki.debian.org/DebianScience
http://blends.alioth.debian.org/science/tasks/meteorology
Alastair McKinstry < mckinstry@debian.org>
Ireland's High-Performance Computing Centre http://www.ichec.ie
```

```
Debian Science: http://wiki.debian.org/DebianScience
  http://blends.alioth.debian.org/science/tasks/meteorology
  Alastair McKinstry < mckinstry@debian.org>
  Ireland's High-Performance Computing Centre http://www.ichec.ie
Debian GIS: http://wiki.debian.org/DebianGis
  http://blends.alioth.debian.org/gis/tasks/
  Francesco P. Lovergine" < frankie@debian.org>
```



Workstation - Geographic Information Systems (GIS) workstation

not restricted to OpenStreetMap data only.

This task sets up your system to be a GIS workstation to process geographical information and make maps.

Debian packages which are dealing with geographical information to be presented for the web on so called map tile servers. These are pretty useful when trying to setup an OpenStreetMap tile server but

Additional benefits (in examples)

- Longevity
- Reproducibility

Longevity

XGKS: Graphical Kernel System for the X Window System

http://xgks.sourceforge.net

- developed within unidata.ucar.edu
- now considered legacy: 2.6 in 2000, 2.6.2 release in 2005-02-03
- still used by
 - Ferret: http://www.ferret.noaa.gov
 - UV-CDAT: http://uvcdat.llnl.gov from LLNL, DOE
- and is still maintained in Debian

Longevity

Recommendations

- do not under-estimate the number of your users
- prepare your software for the after-life:
 - clean legal terms
 - commented code
 - reliable build infrastructure

http://archive.debian.org - archive of ever existed releases

http://archive.debian.org – archive of ever existed releases Example: state of a development box in 2003:

debootstrap --arch=i386 --include=build-essential,emacs20,netcdf3 \
 potato /tmp/potato http://archive.debian.org/debian

http://archive.debian.org – archive of ever existed releases Example: state of a development box in 2003:

```
debootstrap --arch=i386 --include=build-essential,emacs20,netcdf3 \
   potato /tmp/potato http://archive.debian.org/debian
```

■ schroot — enter into the environment upon your convenience

```
> cat /etc/schroot/chroot.d/lenny-20080407
[potato]
description=Antique Debian from 2003
directory=/tmp/potato
users=YOURLOGIN
> schroot -c potato
```

■ http://archive.debian.org - archive of ever existed releases Example: state of a development box in 2003:

```
debootstrap --arch=i386 --include=build-essential,emacs20,netcdf3 \
   potato /tmp/potato http://archive.debian.org/debian
```

schroot – enter into the environment upon your convenience

```
> cat /etc/schroot/chroot.d/lenny-20080407
[potato]
description=Antique Debian from 2003
directory=/tmp/potato
users=YOURI.OGTN
> schroot -c potato
```

http://snapshots.debian.org - snapshots of the archive

```
debootstrap --arch=i386 --include build-essential,libnetcdf-dev \
  testing /tmp/testing \
  http://snapshot.debian.org/archive/debian/20080407T000000Z/
```

Take home

If you care about open, accessible and reproducible science:

Get all software you care about into Debian!

Thanks

http://www.debian.org http://neuro.debian.net

Yaroslav O. Halchenko

yoh@debian.org Michael Hanke mih@debian.org

about the slides:

http://neuro.debian.net/#publications

copyright © 2013 Yaroslav O. Halchenko & Michael Hanke, front/end slide style inspired by Stefano Zacchiroli CC BY-SA 3.0 — Creative Commons Attribution-ShareAlike 3.0

apt-get install -t audience [-a microphone] questions

- ho pprox 18'704 (main), 222 (contrib), 506 (non-free) source packages
- ≈ 38'537 (main), 141 (contrib), 289 (non-free) binary packages
- largest n. of ports among mainstream distros (11 official, 11 unofficial)
 - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)

- ho pprox 18'704 (main), 222 (contrib), 506 (non-free) source packages
- ho pprox 38'537 (main), 141 (contrib), 289 (non-free) binary packages
- largest n. of ports among mainstream distros (11 official, 11 unofficial)
 - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
 - The latest stable release: 6.0 Squeeze, February 6th 2011
 - Upcoming stable release: 7.0 Wheezy, TBA 2013
 - Released \approx 2-3 years
 - Security support for 3 years
 - Upgradable

- ho pprox 18'704 (main), 222 (contrib), 506 (non-free) source packages
- ho pprox 38'537 (main), 141 (contrib), 289 (non-free) binary packages
- largest n. of ports among mainstream distros (11 official, 11 unofficial)
 - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
- 2 1/2 rolling releases: testing, unstable (sid), experimental

Debian, 19 years old: > 100 derivatives

- ho pprox 18'704 (main), 222 (contrib), 506 (non-free) source packages
- ho pprox 38'537 (main), 141 (contrib), 289 (non-free) binary packages
- largest n. of ports among mainstream distros (11 official, 11 unofficial)
 - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
- 2 1/2 rolling releases: testing, unstable (sid), experimental
- > 130 "derivative" distributions (source: distrowatch.com)

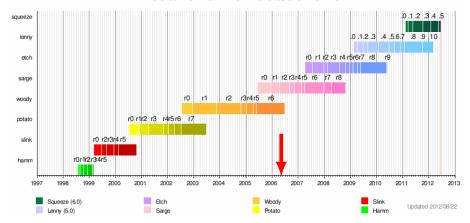
- ho pprox 18'704 (main), 222 (contrib), 506 (non-free) source packages
- ho pprox 38'537 (main), 141 (contrib), 289 (non-free) binary packages
- largest n. of ports among mainstream distros (11 official, 11 unofficial)
 - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
- 2 1/2 rolling releases: testing, unstable (sid), experimental
- > 130 "derivative" distributions (source: distrowatch.com)
- \approx 900 active DDs + \approx 200 DMs + thousands other contributors

Debian: the Universal OS

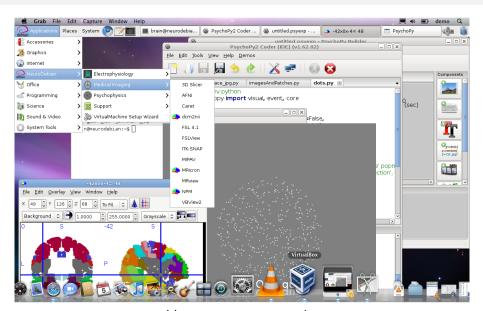
- ho pprox 18'704 (main), 222 (contrib), 506 (non-free) source packages
- ≈ 38'537 (main), 141 (contrib), 289 (non-free) binary packages
- largest n. of ports among mainstream distros (11 official, 11 unofficial)
 - 2 non-Linux ports: GNU/kFreeBSD + (unofficial Hurd)
- 12 stable releases
- 2 1/2 rolling releases: testing, unstable (sid), experimental
- > 130 "derivative" distributions (source: distrowatch.com)
- \sim 900 active DDs + \approx 200 DMs + thousands other contributors



Debian GNU/Linux release timeline



NeuroDebian VM in VirtualBox



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





Journal Info

Neuroscience runs on GNU/Linux

Michael Hanke 1, 2, 3* and Yaroslav O. Halchenko 1, 2

¹ Center for Cognitive Neuroscience, Dartmouth College, USA

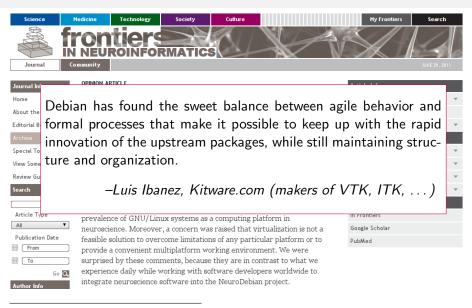
Full Text | Provisional PDF | A A | 🐧 | 🔯 | 🗺 | Share this

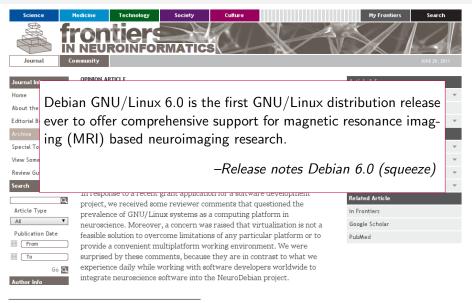
- Department of Psychological and Brain Sciences, Dartmouth College, USA
- 3 Department of Experimental Psychology, Otto-von-Guericke-University, Germany

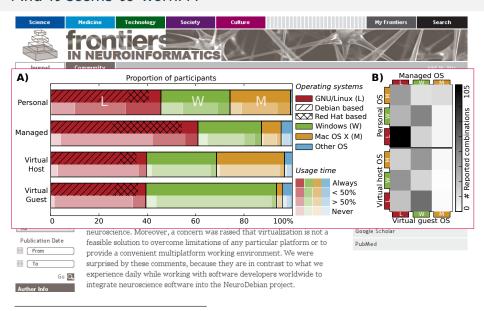
In response to a recent grant application for a software development project, we received some reviewer comments that questioned the prevalence of GNU/Linux systems as a computing platform in neuroscience. Moreover, a concern was raised that virtualization is not a feasible solution to overcome limitations of any particular platform or to provide a convenient multiplatform working environment. We were surprised by these comments, because they are in contrast to what we experience daily while working with software developers worldwide to integrate neuroscience software into the NeuroDebian project.

	 _
Article Info	
Full Text	~
Provisional PDF	
Export Citation	~
The Authors in	
Frontiers	*
Google	~
Google Scholar	~
PubMed	~
Related Article	
in Frontiers	
Google Scholar	
PubMed	

http://www.frontiersin.org/Neuroinformatics/10.3389/fninf.2011.00008/full

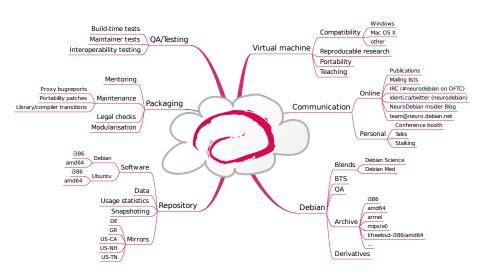






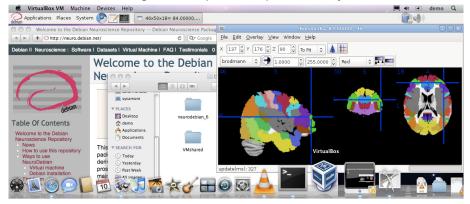
http://www.frontiersin.org/Neuroinformatics/10.3389/fninf.2011.00008/full

What is **NeuroDebian**?



But, but my true love is a fruit and I'm married to monster!

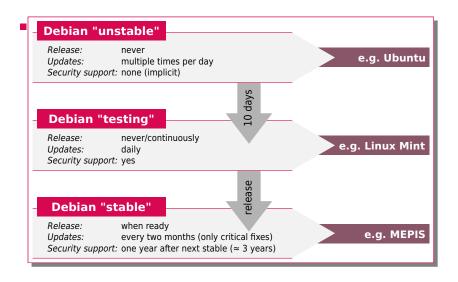
- NeuroDebian virtual machine (32/64bit, multi-core)
- Most convenient solution for Mac OS X, Windows
- Base image with setup wizzard, fully functional within minutes
- Great for teaching, workshops, development, analysis



How does software benefit from Debian?

- Extended reach
 - one stable release, two rolling "release" flavors
 - ≈130 derivative distributions (distrowatch.org)

How does software benefit from Debian?



How does software benefit from Debian?

- Extended reach
 - one <u>stable</u> release, two rolling "release" flavors
 - ightharpoonup pprox 130 derivative distributions (distrowatch.org)
- Mutual awareness
 - Explicitly documented dependencies
 - Synchronized transitions
- Less maintenance work through modularity
 - 3rd-party software in dedicated packages maintained by someone else
- Continuous integration testing
 - 13 hardware architectures
 - Three kernels
 - Continuous automated testing for
 - Build success
 - Clean installation/de-installation, Availability of dependencies
 - Policy compliance
 - Package conflicts

But I only care about Ubuntu!

No, you don't!

- Most software we care about comes (almost) 1:1 from Debian (SciPy, VTK, ITK, . . .)
- No LTS for neuroscience (NumPy only since 10.04)

Go Debian!

- Developers: Get it right in Debian, have it work in Debian/Ubuntu/Mint/aptosid/Mepis/...(at no additional cost)
- Users: Stable release with 3-4 years support for all software
- Scientists: Want your research tool to be found and used? Include it in the largest software archive in the world.

NeuroDebian world-map

