r2u

CRAN as Ubuntu Binaries

Dirk Eddelbuettel

CANSSI Ontario Statistical Software Conference

10 Nov 2022

https://dirk.eddelbuettel.com/papers/canssi_nov2022_r2u.pdf

ILLINOIS

My day job

Principal Software Engineer, TileDB

- building a *universal* data engine for data science, bioinformatics, geospatial, ...

(Adjunct) Clinical Professor, University of Illinois

- teaching STAT447 'Data Science Programming Methods'

Open Source Work

- Debian developer since 1995, currently maintaining about 185 packages
- R package author since 2003, author or maintainer of well over 60 CRAN packages
- R Foundation Board Member; JSS Associate Editor
- Rocker Project co-founder: Docker for R, including official 'r-base' image

My day job

Principal Software Engineer, TileDB

- building a *universal* data engine for data science, bioinformatics, geospatial, ...

(Adjunct) Clinical Professor, University of Illinois

- teaching STAT447 'Data Science Programming Methods'

Open Source Work

- Debian developer since 1995, currently maintaining about 185 packages
- R package author since 2003, author or maintainer of well over 60 CRAN packages
- R Foundation Board Member; JSS Associate Editor
- Rocker Project co-founder: Docker for R, including official 'r-base' image

Status Quo: CRAN and Binaries

Windows: Child's Play!

macOS: Elegant if you
do as you're told

Linux: Ah, yes, ..., packages would be nice to have

CRAN Binaries Across Operating Systems

Windows works pretty well for binaries (but building from *source* can be a pain having to first install `Rtools`, getting other source dependencies, lack of package manager so some flee for Conda ...)

macOS works pretty well for binaries from what one hears (modulo edge uses with OpenMP, or Fortran, or at times building from source, some use `brew` or Conda)

Linux Tower of Babel: some distros have some binary packages of some versions, but users do not always know, installation from sources works "for those who know", can be time confusing, various tricks (`ccache`, c2d4u, ...), some flee for Conda

Today's talk is about getting the Linux use case to the ease of the prior two

R2U

Prior Work on CRAN into Debian / Ubuntu

cran2deb v1 First fully automated conversion of CRAN packages into Debian binaries (via Perl) in early 2000s by Albrecht Gebhard; building on this David Vernazobres, Albrecht Gebhard, Stefan Moeller, Dirk Eddelbuettel had working system with a few thousand binaries (see useR! 2007 talk)

cran2deb v2 Excellent GSoC work by Charles Blundell with a full R rewrite, ~ 6k or 7k packages, I ran it for a ~ year til machine died (see useR! 2009 talk)

debian-r During his PhD studies, Don Armstrong did full CRAN and BioC builds til his machine died (2015 personal reference)

c2d4u Extending the cran2deb work, Michael Rutter with new approach using Launchpad, still active after 10+ years, now ~ 5k packages (useR! 2011 talk)

r-builders R Consortium Grant for Michael Rutter, Don Armstrong, Gabor Csardi, Dirk Eddelbuettel: Grant received, later returned as no liftoff

Conclusions

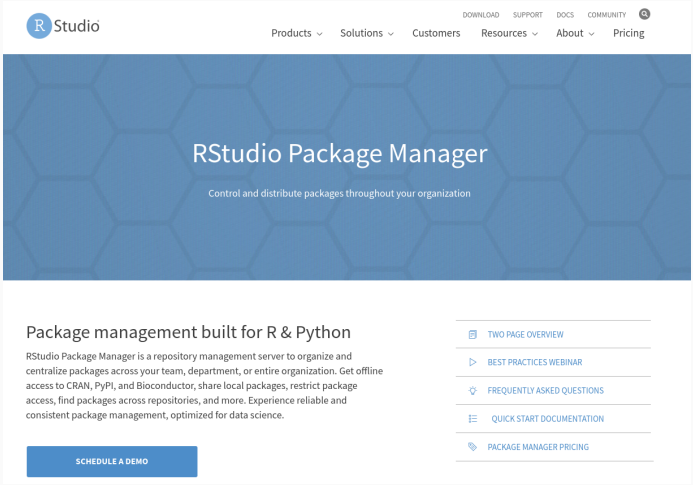
- Building all of CRAN from source is *hard* and *a lot of work*
- It can be done (*c.f.* Don Armstrong), and *it is being done*
  - Iñaki Ucar for Fedora (leaning on Fedora build infra)
  - Detlef Steuer for OpenSUSE (leaning on OBS)
  - Michael Rutter for Ubuntu, at least partially (leaning on Launchpad)
- It seems to work better if we can rely on systems that build packages
- Question: How do we fill the gap of a *complete* Ubuntu (and/or Debian) solution?

Not Really Enough Time in Today's Short Talk ...

- What is a .deb package?
- What is inside a .deb file?
- What does `R CMD INSTALL --build .`
- Simple tree of files, *no metadata*

RSPM (RStudio Package Manager)

RSPM


Studio
DOWNLOAD
SUPPORT
DOCS
COMMUNITY
Products
Solutions
Customers
Resources
About
Pricing
RStudio Package Manager
Control and distribute packages throughout your organization
Package management built for R & Python
RStudio Package Manager is a repository management server to organize and centralize packages across your team, department, or entire organization. Get offline access to CRAN, PyPI, and Bioconductor, share local packages, restrict package access, find packages across repositories, and more. Experience reliable and consistent package management, optimized for data science.
SCHEDULE A DEMO
TWO PAGE OVERVIEW
BEST PRACTICES WEBINAR
FREQUENTLY ASKED QUESTIONS
QUICK START DOCUMENTATION
PACKAGE MANAGER PRICING

RSPM Example (on my OS + R Version)

```
edd@rob:~$ wget -q https://packagemanager.rstudio.com/cran/__linux__/jammy/latest/src/contrib/RcppNLoptExa
mple_0.0.1.tar.gz?r_version=4.2 -O RcppNLoptExample_0.0.1.tar.gz
edd@rob:~$ tar tvzf RcppNLoptExample_0.0.1.tar.gz
-rw-rw-r-- builder/builder 973 2022-07-12 12:57 RcppNLoptExample/DESCRIPTION
-rw-rw-r-- builder/builder  76 2022-07-12 12:57 RcppNLoptExample/INDEX
drwxrwxr-x builder/builder   0 2022-07-12 12:57 RcppNLoptExample/Meta/
-rw-rw-r-- builder/builder 278 2022-07-12 12:57 RcppNLoptExample/Meta/Rd.rds
-rw-rw-r-- builder/builder 121 2022-07-12 12:57 RcppNLoptExample/Meta/features.rds
-rw-rw-r-- builder/builder 288 2022-07-12 12:57 RcppNLoptExample/Meta/hsearch.rds
-rw-rw-r-- builder/builder 142 2022-07-12 12:57 RcppNLoptExample/Meta/links.rds
-rw-rw-r-- builder/builder 333 2022-07-12 12:57 RcppNLoptExample/Meta/nsInfo.rds
-rw-rw-r-- builder/builder 1012 2022-07-12 12:57 RcppNLoptExample/Meta/package.rds
-rw-rw-r-- builder/builder  134 2022-07-12 12:57 RcppNLoptExample/NAMESPACE
-rw-rw-r-- builder/builder  385 2022-07-12 12:57 RcppNLoptExample/NEWS.Rd
drwxrwxr-x builder/builder    0 2022-07-12 12:57 RcppNLoptExample/R/
-rw-rw-r-- builder/builder 1058 2022-07-12 12:57 RcppNLoptExample/R/RcppNLoptExample
-rw-rw-r-- builder/builder 1366 2022-07-12 12:57 RcppNLoptExample/R/RcppNLoptExample.rdb
-rw-rw-r-- builder/builder  262 2022-07-12 12:57 RcppNLoptExample/R/RcppNLoptExample.rdx
drwxrwxr-x builder/builder    0 2022-07-12 12:57 RcppNLoptExample/help/
-rw-rw-r-- builder/builder   46 2022-07-12 12:57 RcppNLoptExample/help/AnIndex
-rw-rw-r-- builder/builder 1484 2022-07-12 12:57 RcppNLoptExample/help/RcppNLoptExample.rdb
-rw-rw-r-- builder/builder  171 2022-07-12 12:57 RcppNLoptExample/help/RcppNLoptExample.rdx
-rw-rw-r-- builder/builder  107 2022-07-12 12:57 RcppNLoptExample/help/aliases.rds
-rw-rw-r-- builder/builder  156 2022-07-12 12:57 RcppNLoptExample/help/paths.rds
drwxrwxr-x builder/builder    0 2022-07-12 12:57 RcppNLoptExample/html/
-rw-rw-r-- builder/builder 1174 2022-07-12 12:57 RcppNLoptExample/html/00Index.html
-rw-rw-r-- builder/builder 1735 2022-07-12 12:57 RcppNLoptExample/html/R.css
drwxrwxr-x builder/builder    0 2022-07-12 12:57 RcppNLoptExample/libs/
-rwxrwxr-x builder/builder 1006520 2022-07-12 12:57 RcppNLoptExample/libs/RcppNLoptExample.so
edd@rob:~$
```

We can download a tar.gz

No metadata whatsoever!

So this is like `R CMD INSTALL --build` !!

About RSPM

Positive Aspects

- Windows + macOS + various Linux distros, also multiple R versions, even Python
- 'just the binary' – what `R CMD INSTALL --build` creates

Less Positive Aspects

- no (real, automatic, full) system dependencies
- no system management integration (on OSs that have it)
- some packages are source and *still* need building
- no Debian, no arm64

Overall

- not bad at all, some warts (esp. for Linux users) notwithstanding

The Key Insight behind r2u

What makes it sing

- use build artifacts – which RSPM gives us – as input in `.deb` build step
- use distro package tools for *proper* integration (given parameterization)
- "cheaply, quickly, reliably" create packages *with full system dependencies*
- including from source where needed using the standard full-build way

Enter r2u

- operational and up since May
- Internet2 connected mirror added in September
- currently serving between 5k and 10k binaries per (business) day

Demo

Time permitting...

See website for demos, and ability to run ‘in the browser’ via gitpod.io.

Usage

Setup Script

Five Very Simple Steps (And Four Are Optional)

- Update `apt` indices, install `wget` and certs
- Add `r2u` repo and keys for secure apt
- Update R itself to ensure it is current
- Add pinning to ensure proper package sorting
- Add `bspm` for easier installation from R

```
#!
## Note that you need to run this as root

## First: update apt and get wget to fetch keys
apt update -qq
apt install --yes --no-install-recommends wget ca-certificates

## Second: add the CRAN apt repo and key -- here we now use the mirror
wget -q -O- https://eddelbuettel.github.io/r2u/assets/dirk_eddelbuettel_key.asc \
    | tee -a /etc/apt/trusted.gpg.d/cranapt_key.asc
#echo "deb [arch=amd64] https://ltl.rc.eddelbuettel.com/cranapt jammy main" > /etc/apt/sources.list.d/cranapt.list
echo "deb [arch=amd64] https://r2u.stat.illinois.edu/ubuntu jammy main" > /etc/apt/sources.list.d/cranapt.list
apt update

## Third: update current R (jammy has 4.1.2 but we want 4.2.2 or newer)
wget -q -O- https://cloud.r-project.org/bin/linux/ubuntu/marutter_pubkey.asc \
    | tee -a /etc/apt/trusted.gpg.d/cran_ubuntu_key.asc
echo "deb [arch=amd64] https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/" > /etc/apt/sources.list.d/cran_r.list

## Fourth: add pinning to ensure package sorting
echo "Package: *" > /etc/apt/preferences.d/99cranapt
echo "Pin: release o=CRAN-Apt Project" >> /etc/apt/preferences.d/99cranapt_
echo "Pin: release l=CRAN-Apt Packages" >> /etc/apt/preferences.d/99cranapt
echo "Pin-Priority: 700"  >> /etc/apt/preferences.d/99cranapt

## Fifth: install bspm and enable it
##
## If needed (in bare container, say) install python tools for bspm and R itself
#DEBIAN_FRONTEND=noninteractive apt install --yes --no-install-recommends python3-dbus python3-apt r-base-core
##
Rscript -e 'install.packages("bspm")'
export RHOME=$(R RHOME)
echo "suppressMessages(bspm::enable())" >> ${RHOME}/etc/Rprofile.site
## Giving bspm sudo right used to be required but no longer is under current bspm versions
#echo "options(bspm.sudo=TRUE)" >> ${RHOME}/etc/Rprofile.site
```

See the script directory and either run the script (or steps one by one) on any Ubuntu system.

Docker Containers

Use the `eddelbuettel/r2u` containers

- For Ubuntu 20.04 and 22.04
- Likely ‘soon’ part of Rocker too

Also For Example Available for Use at Gitpod.io

- See the `r2u` README.md with a link to try

In GitHub Actions (and other CI systems!)

Easy as r-ci now uses r2u

- download `run.sh`
- bootstrap to add repos and `r2u`
- add dependencies via `install_dep` (with suggests via `install_all`)
- build package and run tests

That is all – see my repos for examples.

"Fast, Cheap, Reliable – pick any three!"

Key portion from one of many identical `ci.yaml` files:

```
steps:
  - uses: actions/checkout@v3

  - name: Get Script
    run: curl -OLs https://eddelbuettel.github.io/r-ci/run.sh\
           && chmod 0755 run.sh

  - name: Bootstrap
    run: ./run.sh bootstrap

  - name: Dependencies
    run: ./run.sh install_deps

  - name: Test
    run: ./run.sh run_tests
```

Summary

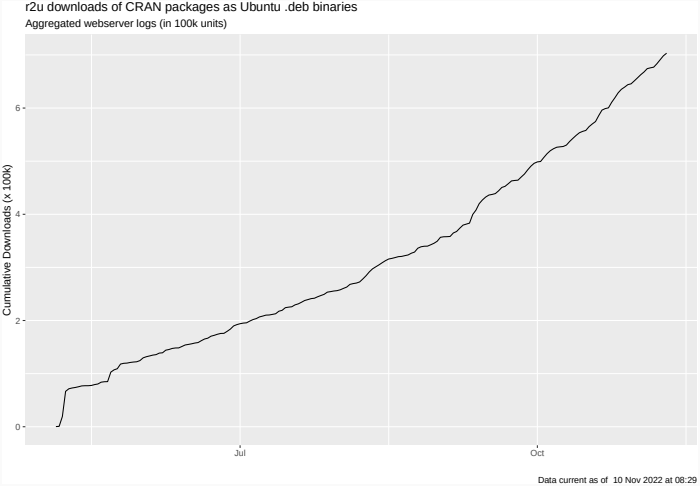
Binary CRAN Packages Rock

- Now you can rock them whereever Ubuntu runs
- Your server, and if so inclined, your desktop / laptop
- Your cloud instance at AWS, GCS, Azure, ...
- Your continuous integration integration action runner
- You name it: Linux and Ubuntu are fairly universal
- **r2u** gives you all of CRAN at an instant with full dependency resolution

An Optional Extra Cherry on Top

- The `bspm` package by Iñaki is an extra cherry on top
- It `trace()`s the `install.packages()` function
- And connects it to `apt` with proper 'translation'
- So `install.packages("ggplot2")` becomes ...
- ... `sudo apt install r-cran-ggplot2`
- Very convenient here as we do not have to explain `apt` naming
- But `r2u` (as a repository) works with or without it

r2u Use


r2u downloads of CRAN packages as Ubuntu .deb binaries
Aggregated webserver logs (in 100k units)
Cumulative Downloads (x 100k)
0
2
4
6
Jul
Oct
Data current as of 10 Nov 2022 at 08:29

That's It!

Thanks to

- R (package) authors for creating something wonderful in the commons
- The CRAN team for all they do making it *reliably* accessible
- Albrecht, David, Stefan, Charles, Don, Michael, ... for all the earlier work
- RSudio / Posit for RSPM / PPM, and Iñaki for BSPM
- Statistics at the U of Illinois Urbana-Champaign for hosting r2u
- Rami Dass for invaluable and patient help in making that happen
- Pixabay for providing the 'free use / no attribution needed' images
- My GitHub sponsors for all the coffee money

And see https://eddelbuettel.github.io/r2u/ for **r2u**