

NHS-R community: Supporting users and  
publishing open source in UK health and social  
care.

Chris Beeley, Nottinghamshire Healthcare NHS Trust

8th September 2022

# What is NHS-R?

- ▶ Analysts and data scientists in health and social care
- ▶ Using open source tools (mainly R!)
- ▶ We share code (everything defaults to OPEN)
- ▶ Cooperate across organisational boundaries
- ▶ We cooperate across international boundaries
- ▶ We love beginners
- ▶ We make mistakes and learn together

## Why does NHS-R exist?

- ▶ Training
- ▶ Community
- ▶ Policy/ good practice
- ▶ Building tools for everyone to use- WAGOLL
- ▶ Lead by example- share, be inclusive, and be attractive to beginners

# Open source

- ▶ Natural to assume that the NHS is a collaborative enterprise
- ▶ Until recently there were no mechanisms to help us cooperate at all
- ▶ Everything is done over and over again, duplicatively, on proprietary DBs
- ▶ In the recent past:
  - ▶ No incentive to share
  - ▶ Many barriers to sharing

## Who does share?

- ▶ Facebook
- ▶ Microsoft
- ▶ IBM
- ▶ Samsung

# Let's talk about Prophet

- ▶ What is Prophet?
- ▶ Why did Facebook give away Prophet (MIT licence)?
- ▶ I really have no idea, but let's speculate anyway
  - ▶ Developers like to work in the open so it helps with recruitment and retention
  - ▶ They want people to use it so they can get better suggestions about how to improve it
  - ▶ They want people to improve it by submitting pull requests (several at the time of writing)
  - ▶ They don't want the hassle of selling it!

# Why would we share?

- ▶ The four freedoms
  - ▶ Use
  - ▶ Study
  - ▶ Share
  - ▶ Improve

# How are we doing?

- ▶ Some orgs are still fighting about use!
- ▶ Many fighting tooth and nail over learning and transparency
  - ▶ Why?
- ▶ Building? Largely a pipe dream
  - ▶ Why?
- ▶ Contributing? No!
  - ▶ Why? *The Daily Mail test*



# What can NHS-R do about this?

- ▶ Permission
- ▶ Visibility and community
- ▶ Skills
  - ▶ Basic- git, GitHub, issues, pull requests...
  - ▶ Advanced- building reusable stuff is hard
  - ▶ Advanced advanced- building reusable stuff *together* is really hard!

## Building tools

- ▶ Two tools today
- ▶ One for every provider trust in the country- simple, RAP compliant, useful
- ▶ One to show the potential for open source data science

# NHSRplotthedots

- ▶ SPCs and Making Data Count
- ▶ Done up and down the country with Excel templates
  - ▶ Not RAP compliant
  - ▶ ... what is RAP?
- ▶ We discussed the idea of an R implementation of this for a long time
- ▶ In the end someone just got on with it
- ▶ Someone from the community shared code they'd already written
- ▶ Group effort to expand and improve
- ▶ Now on CRAN with >1500 downloads

# Community roles

- ▶ Methodology
- ▶ Documentation
- ▶ User requirements and testing
- ▶ Code and repo management
- ▶ Wider community engagement

## Why it succeeded

- ▶ Real business need
- ▶ Grew organically from users deploying the package
- ▶ Community roles
- ▶ CRAN package == Good for security departments in NHS
- ▶ Promoted by MDC
- ▶ User engagement, teamwork
  - ▶ A truly collaborative, high quality product
  - ▶ Funding: £0

## Text mining of patient experience data

- ▶ One year funded project (NHSE)
- ▶ Now funded for another year
- ▶ NHS-R funded an R wrapper to the Python code in the project
- ▶ Python code, Shiny dashboard
  - ▶ Ship the code as MIT and do a managed deployment on RStudio Connect

# Success!?

- ▶ The code works and the evaluation was a success
- ▶ There is a real and growing need for text analytics in healthcare
  - ▶ DMs open 😊
- ▶ Judged by my own standards the project is not a success (yet!)
  - ▶ Nobody outside the team has ever contributed to the code
  - ▶ Nobody outside the team has even run the code

# What are the barriers

- ▶ Code is complex
  - ▶ The whole pipeline is in two languages
- ▶ Many of the pilot sites are locked into contracts with other providers
  - ▶ Including collection and collation
- ▶ BI systems are not usually in R
  - ▶ (even ours aren't- except patient experience because I wrote it)
- ▶ It's too cheap!



# What can we do?

- ▶ Build tools, not pipelines
  - ▶ Do one thing well
- ▶ Start charging for open source (?!)
- ▶ Get in the weeds and deploy it trust by trust (?!)
- ▶ Listen to users and developers

# Summary

- ▶ Open source means efficiency, transparency, and building the stuff we *actually want*
- ▶ Open source is hard. Using the tools is hard and writing the code is hard
- ▶ Having permission to share (never mind about collaboratively build) is hard
- ▶ NHS-R exists to give people the permission, skills, and sense of belonging necessary to build software together

# The projects

- ▶ NHSRplotthedots is a testament to the power of community
  - ▶ Does one thing, everybody wants it, easy to deploy, created spontaneously
  - ▶ WAGOLL
- ▶ Text mining is a problem to be solved
  - ▶ I want to ship the code
  - ▶ I want others to contribute
  - ▶ I want someone else to host it, even
- ▶ Can we build bigger than NHSRplotthedots?
- ▶ Can we build smaller than patient experience text mining?