

NHS-R Community Handbook

NHS-R Community

Last compiled 2022-05-30

Contents

| | |
|------------------------------------|-----------|
| | 5 |
| 1 Purpose | 7 |
| 2 Open code | 9 |
| 2.1 Introduction | 9 |
| 2.2 References | 9 |
| 2.3 Specific challenges | 10 |
| 3 Statement on using tools | 11 |
| 4 NHS-R vision | 13 |
| 4.1 Using R in research | 13 |
| 4.2 Training | 14 |
| 4.3 Development | 14 |
| 4.4 Events | 14 |
| 4.5 Appendix A: workshop | 14 |

In this book we've compiled a set of resources for anyone using R within the NHS and beyond. These are taken directly from resources we created within our GitHub repos. This handbook is updated once a day to pull in any changes that have been made in those repos.

If you'd like to see additional resources included here, feel free to open an issue!

Chapter 1

Purpose

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Chapter 2

Open code

2.1 Introduction

There are number of organisations and publications advocating open coding in its various guises within the NHS and more generally across the public sector and academia. This page is intended as a place to collate pertinent literature and proposed/tried approaches to implementation of open code policies in situations in the NHS from individual teams up to entire organisations.

There is also a significant crossover between open code and software engineering best practice, which means that they are often (always?) promoted together as an effective means to improve reproducibility,

2.1.1 Glossary (DRAFT)

- **Reproducible Analytics Pipelines (RAP):** analytics processes developed in open source programming languages and adhering to software engineering best practices to allow for reproducing analyses with very little effort.
- **Repo/Repository:** a set of files organised in a project for a specific purpose (such as statements-on-tools itself), containing code or documentation under version control.

2.2 References

The following is a list of some of the material available online discussing and supporting open coding and software engineering approaches to code development in the NHS:

- The Goldacre Report - a systematic and far-reaching report, written on behalf of the Department of Health, advocating for open coding and re-

producible analytical pipelines (RAP) in the NHS,

- Office for Statistics Regulation – Overcoming Barriers to Adoption of RAP - a report written in support of RAP for adoption by all government departments doing analytics, describing the challenges and recommending solutions to address those challenges covering both organisational, team-level and individual barriers,
- NHSX Open Source Policy (currently in DRAFT) - a comprehensive description of why and how open source should be implemented in the NHS, including statements about best practice and a checklist for open sourcing code,
- NHSD RAP Community of Practice - a repository containing a wealth of material pertaining to setting up and running RAP, including open source - based on their internal work to implement RAP in various NHSD analytical teams.

2.3 Specific challenges

2.3.1 Opening code without exposing data/‘secrets’

This issue is often raised (and rightly) when discussing open sourcing of code - how do you ensure that personal clinical or other sensitive data are not shared? In practice there are a number of approaches to this (see the NHSX Open Source Policy for more practical advice on this) including setting up repos with automated checking for datasets and secrets (such as API keys) in order that risk of unintentional disclosure is minimised.

2.3.2 Opening code without giving away commercial/business/clinical proprietary information (and IP)

Many members of the NHS-R community work in teams or individually without any formal training or support to enable them to determine whether there would be significant ramifications of publishing particular code, especially where business processes might be encoded in the project, or where there might be some IP issue. This makes the prospect of open sourcing rather daunting.

2.3.3 Coding in the open

The NHSX Open Source Policy recommends that all development/analytics work done in the NHS be coded in the open unless there is good reason not to. However it also states that “an internal code review should be conducted for all open source projects before publication”, which is at odds with coding in the open.

Chapter 3

Statement on using tools

This is an evolving document which describes how and why to use R and other data science tools and to share and reuse code safely in health and social care settings.

The scope and content are expanding all the time as the community collaboratively produces a definitive statement of the “NHS-R way”.

Please file issues, make pull requests, and get involved, we’re very happy from hear from friends from inside and outside of NHS-R.

Read it here <https://nhs-r-community.github.io/statements-on-tools/>

Chapter 4

NHS-R vision

4.1 Using R in research

A mistake in the operating room can threaten the life of one patient; a mistake in statistical analysis or interpretation can lead to hundreds of early deaths. So it is perhaps odd that, while we allow a doctor to conduct surgery only after years of training, we give SPSS® (SPSS, Chicago, IL) to almost anyone. Moreover, whilst only a surgeon would comment on surgical technique, it seems that anybody, regardless of statistical training, feels confident about commenting on statistical data. <https://www.nature.com/articles/ncpuro0294>

The NHS, as one of the largest hospital and healthcare systems, is a world leader in research. Research and evaluation are carried out as funded projects as well as unfunded audits/ evaluation. Both of which often require statistics- the analysis often being done in SPSS/ SAS or Excel. These methods can produce flawed analyses which, moreover, are not reproducible.

Many trusts do not employ statistics experts and will only be able to get statistical help on funded work by buying in time from academic/external statisticians. This means that the pilot work that clinicians do prior to applying for large grants can often be flawed, or promising work ends up not being completed and the grants never awarded because they didn't have the statistics expertise.

While we would not expect clinicians to become expert coders, the NHS-R community should work to develop and deliver training that would help clinicians to be able to use R, including the development of training specifically for those with a clinical/ non coding background. This training needs to include R for statistics as well as the more commonly included data wrangling and visualisation.

Better collaboration between R users working in academia and those in the NHS

would also be beneficial.

4.2 Training

The NHS-R community has developed/ is developing training on introduction to R, Shiny, R Markdown, git, and interactive plotting, amongst other things.

- Where are the gaps in training provision at the current time?
- How can NHS-R maintain/ increase the amount of training that it is able to deliver free to health and social care staff in the NHS in the UK?

4.3 Development

The NHS-R community comprises members with a very wide diversity of job roles and skills. Although there is no one route to being a skilled and useful R developer in health and social care nonetheless the R community could usefully contribute to thinking on how to recruit, train, and develop analysts who use R and other open code approaches to data science/ data analysis.

4.4 Events

NHS-R already has a very successful conference but in the workshop there was a suggestion that there could be another NHS-R event focused on finding problems and solutions to them. There are several ways this could usefully be done, perhaps as a hackathon type approach where the problems are begun on during the event (finished during the event if there is time or worked on afterwards), or more of a problem definition/ alliance building type approach where the actual problems and the people who are interested in solving them are identified during the event and then the actual development is done later

4.5 Appendix A: workshop

The following summarises

4.5.1 How has the NHS-R Community contributed to the system thus far?

The positive contribution of the NHS-R Community was shared by all stakeholders and included the following highlights. For analysts: a safe, trusted, supportive space to learn and share together, a badge of honour, joy, confidence, upskilling, networking, working across organisational boundaries without needing permission. For leaders: signposting to a trusted brand and community that can influence policy which is underpinned by two value systems - NHS and open source. For the wider system: NHS-R has shifted thinking on how to

secure analytic needs in the future and is perhaps the world's first open-source community focused on health and care with admirers across the globe.

4.5.2 What is the need in the system given the change in the health and care landscape?

The system has evolved, and progress has been made since the NHS-R Community was established. Still, some common themes persist regarding the needs of the system; skills gaps, infrastructure needs, better collaborative working and more structured peer learning, and the development of analytical leadership. Specific examples include: methodological training for analysts, setting professional standards, equipping leaders with analytical thinking skills, supporting the use of operations research methods, building links with social care and researchers, quality assurance processes and more. There was consensus that the NHS-R Community should focus on its strengths and not duplicate or drift from this because this may undermine its impact (perhaps because it becomes less relevant to its core members who are flourishing in the freedom of NHS-R). Below is a summary of what the "NHS-R Community" can do to support the system.

4.5.3 What can the NHS-R Community do to support the system?

The following areas were identified. 1. Engage with NHS leaders to help them appreciate the potential of the NHS-R Community as a resource. 2. Work with NHSD/Transformation Directorate to remove barriers/create resources for IT departments to make open source tools readily available for analysts. 3. Provide an 'Ask us' hub where leaders and analysts can refer their questions or issues so that they can get a "grass roots" view from the NHS-R Community on how these might be best addressed. 4. Scaling local solutions nationally and vice-versa 5. Myth busting on "open-source" analytics including addressing security and information governance concerns. 6. Increased collaboration with national bodies such as NHS Transformation Directorate. 7. Set up an NHS Data Science Event (say over 3 to 5 days) for the NHS to identify common problems and develop shared solutions.

4.5.4 We asked a pre-mortem question - imagine the NHS-R Community has died, what led to its demise?

The following were identified.

1. The NHS-R Community was too reliant on volunteers who were unable to sustain their input.
2. The NHS-R Community lost its values and was no longer a brand that was seen as safe, trusted, welcoming, especially to newbies.
3. The NHS-R Community got too pre-occupied with contributing to the centre and so lost touch with grass roots analysts.

4. National data science teams/bodies did not feel as if they had a stake in the NHS-R Community and so disengaged with it and could not see its relevance.
5. R lost out to Python or some other open data science tool.
6. The NHS-R Community did not offer enough “attractors” to analysts (eg wider training, support, development opportunities, etc).
7. The NHS-R Community did not have adequate funds to continue to support it.
8. The NHS-R Community lost its central organising team and so disintegrated.

This should inform our approach to risk over the coming years. We should focus on resilience, being forward thinking and responsive, maintaining our values, broadening the organising team, actively seeking & cultivating new members, and finding funding solutions and partners that can support our activities.