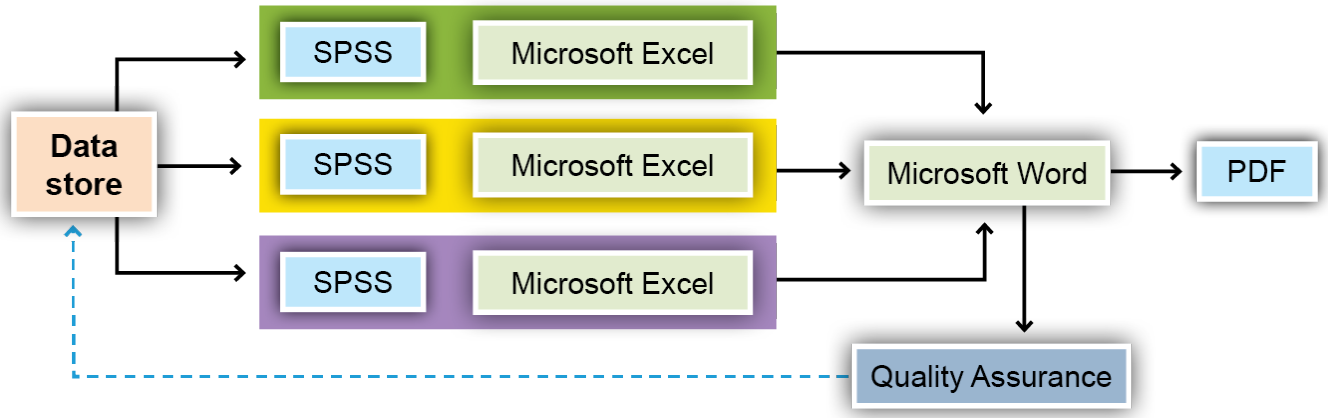
Reproducible Analytical Pipelines (RAP)

Transforming PHI Publications through RAP

Simon Quirk Information Services Division

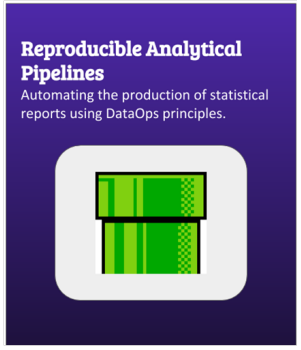
27 January 2020 Slides:https://github.com/simonq01/RAP-Awareness-session-HS-Jan-2020

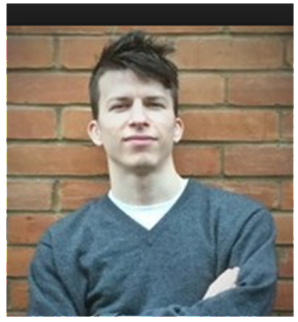
Current publication process



- Complex (many steps between software)
- Prone to error
- Manual, menial tasks carried out by highly skilled people
- Not reproducible or sustainable

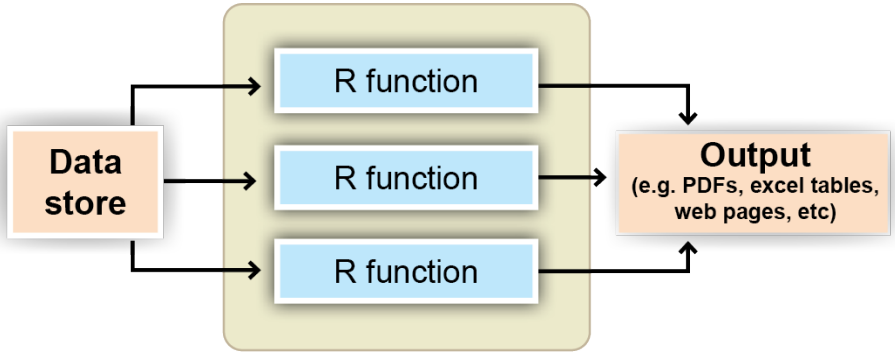
The solution





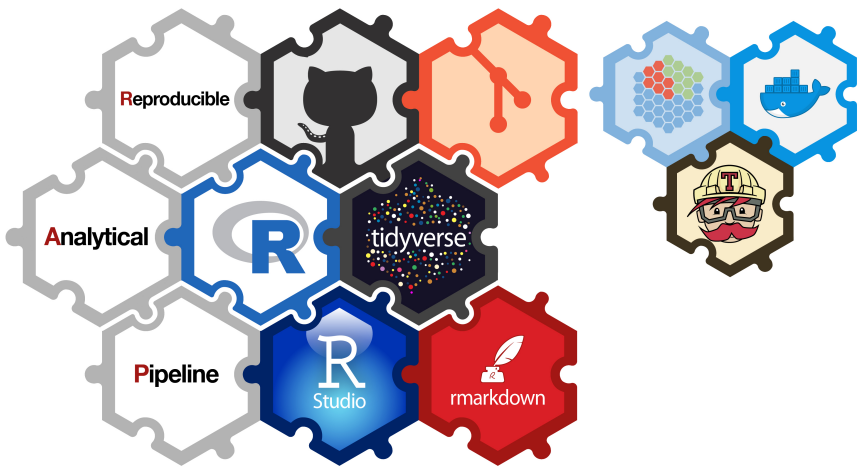
RAP companion

Combined the principles of **reproducible research** with **data science tools and best practice**.

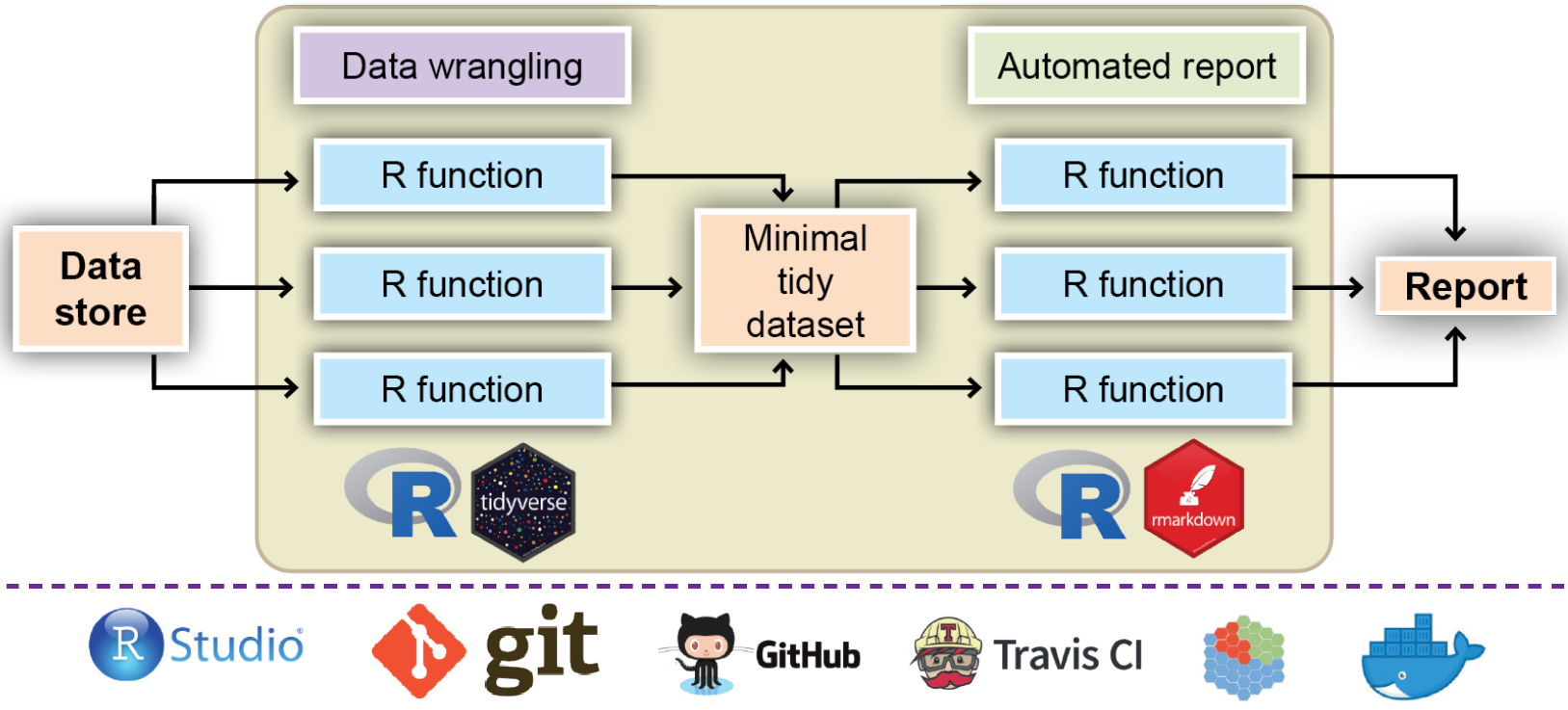


What is RAP?

- No (or few) manual steps = data and output produced using code
- High quality and auditable = version control
- Sustainable = peer review
- "Bells and whistles" = *functions, documenting/testing these functions, package management and computing environment*



What is RAP?



Reproducibility and quality of pipeline ensured via: RStudio project; git for version control; GitHub for code repository and peer review; unit testing of functions; R package management; and computing environment control.

Levels of RAP/automation

Level	Description
1	Ad hoc R code
2	R project
3	R project under version control (VC)
4a	R project under VC and peer reviewed (wrangling)
4b	Replicable report in Rmarkdown (publication)
5	Near RAP (VC, peer review, data quality assurance)
6	Full RAP (as above plus unit testing and documentation)
7	R package

Challenges

- Culture change (peer review and working in the open)
- Senior management support
- New skills for analysts to learn (e.g. R, git)
- Required development time
- Range of data sources and/or unstable production process
- IT (RStudio server and internally hosted code repository)

How to scale RAP in ISD?

- The Transforming Publishing (TPP) team have begun to roll out RAP to other teams in ISD using a buddy system
- One or two members of TPP 'buddy up' with another team to help them create a Reproducible Analytical Pipeline for their publication
- The bulk of the development work is done by the team being 'buddied'; TPP use a light touch approach to provide code reviews; offer assistance with R and Git; advise on timelines; and, more generally, offer guidance wherever it is required
- As a minimum, we recommend teams aim for level 4 (a or b or both) as laid out in our RAP paper

Before the buddying

- Prior to working on RAP, analysts in ISD must attend an introductory Tidyverse training course (run by Jumping Rivers) and a tutorial on using Git and GitHub (run internally by TPP)
- Following this, we ask them to complete an R Skills survey and a short exercise to assess their competence with R
- We also ask them to look at our Toolkit, which contains links to our resources on R, RMarkdown, version control, RAP, shiny and more

Buddying in action

- The first publication to undergo the buddy system is the Scottish Bowel Screening Programme Statistics (SBSS)
- Two analysts created a document detailing the sections of their publication report which required to be automated
- They also created a plan detailing the scripts they required to write and the associated timelines updated fortnightly
- Since the beginning of April, they've converted most of the back end of the publication from SPSS to a version controlled, peer reviewed R project held in a GitHub repository
- Other publications we are buddying with: End of Life Care, Medicines and Mental Health, Cancelled Planned Operations

Interested in RAP?...

- How many reports do your team produce?
- What proportion of time is spent producing reports?
- How much copying and pasting/data movement between software is involved?
- What proportion of your spreadsheet or report contains errors?
- What would the impact of mistakes in production be?
- Could your team create the report if certain team members suddenly left?
- Could you reproduce your publication statistics from 5 years ago?

Contact the Transforming Publishing team
(nss.isdtransformingpublishing@nhs.net)

Thank You

Contact Transforming Publishing:
nss.isdtransformingpublishing@nhs.net

Transforming Publishing web pages

Transforming Publishing on GitHub

PHI on GitHub

RAP companion