

Using Git and GitLab to Empower QC Workflow and Project Management in Clinical Reporting

Kieran Martin, Senior Principal Data Scientist, DSX

Table of contents

- Introduction
- What's Version Control?
- Why might Git be good for QC?
- Our Git Strategy
- Challenges

Introduction

What's version control?

What is Version Control?

- ☐  myawesomeproject.Rmd
- ☐  myawesomeprojectv2.Rmd
- ☐  myawesomeprojectv3.Rmd
- ☐  myawesomeprojectv4.Rmd
- ☐  myawesomeprojectvfinal.Rmd
- ☐  myawesomeprojectvfinalv2.Rmd
- ☐  myawesomeprojectvreallyfinal.Rmd
- ☐  myawesomeprojectusethisone.Rmd
- ☐  myawesomeproject_starthere2024.Rmd
- ☐  myawesomeproject_starthere2024_v2.Rmd
- ☐  myawesomeproject_starthere2024_v3.Rmd

Let's avoid this!

What's Git?



- Developed in **2005** for Linux Kernel
- **Free** and **Open Source**
- 2022 93.87% Stack Overflow Developer Survey respondents reported using it
- Designed to be **fast** and **reliable**

What makes Git Special?





Websites which host git repositories

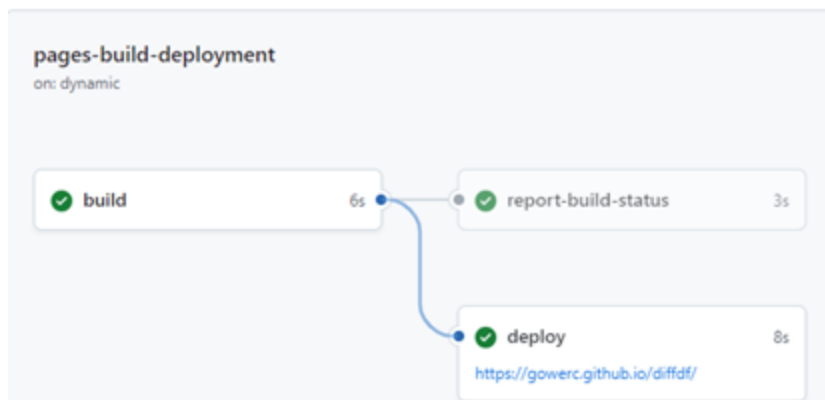
DESCRIPTION	Release v1.1.1 (#131)	3 weeks ago
LICENSE	Minor edits getting package ready for cran	6 years ago
NAMESPACE	Add datetime unit tests (#123)	2 months ago
NEWS.md	Release v1.1.1 (#131)	3 weeks ago
README.md	Release 1.1.0 (#129)	3 weeks ago
_pkgdown.yml	Convert to GHA CI/CD (#90)	6 months ago
cran-comments.md	Release v1.1.1 (#131)	3 weeks ago
difffrproj	left over stuff that I had previous forgot to commit	5 years ago

README License

diffdf

Cran 1.1.1 Downloads 1,702/month Test Coverage 94.75%

diffdf: compares two `data.frame` objects and provides a detailed summary of any differences that were found



So roughly ~ 36% performance improvement which I think is consistent with what we saw last time we tried to make this change.

I will say the implementation is awkward... `data.table` has many weird quirks to it. I managed to get all the tests passing but I wouldn't be surprised if there are edge cases I haven't handled properly yet...



gowerc linked a pull request on Jul 29 that will close this issue

Use `data.table` backend #125

21 Open

kieranmartin commented on Jul 29

Collaborator

So assuming we always change it back to `tibble` before returning the comparison object, the biggest downside of these changes are for users who want to extend `diffdf`. They will probably need to have to interact with `data.table`, and as you say if you are not familiar with it it is basically a completely different language, both in terms of syntax and the behaviour of `data.table` objects when

gowerc requested your review on this pull request.

Use `data.table` backend #125

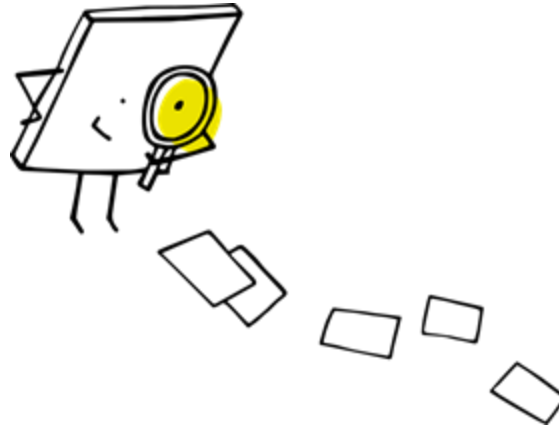
Why might Git be good for QC?

QC Requirements

What?

How?

Who?



Audit: Telling a story

Activity

- Kieran Martin added **Risk** **High** **QC-Method** **DP** labels 1 week ago
- Kieran Martin assigned to **@martik32** 1 week ago
- Kieran Martin created branch **1-dm** to address this issue 1 week ago
- Kieran Martin mentioned in merge request **!1 (merged)** 1 week ago
- Kieran Martin mentioned in commit **c0966c23** 1 week ago
- Kieran Martin closed with commit **c0966c23** 1 week ago

Resolve "dm"

Merged Kieran Martin requested to merge **1-dm** into **devel** 1 week ago

Overview 0 **Commits** 1 **Pipelines** 0 **Changes** 2

resolve #1

parent **845a990e**

Branches > **Branches containing commit**

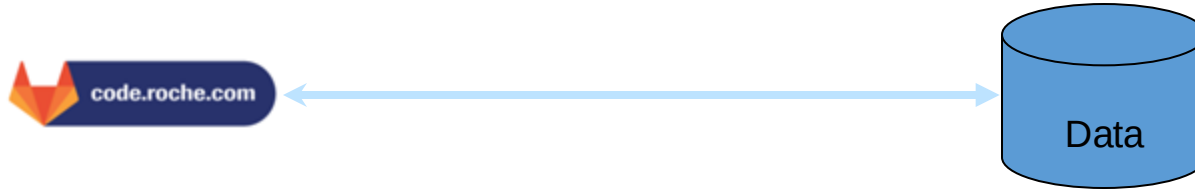
3 merge requests **!7** Fix conflict, **!5** Devel, **!1** Resolve "dm"

Changes 2

Showing **2 changed files** with **0 additions** and **9 deletions**

libraries.R

What makes gitlab so good?



Advanced search is enabled. [View syntax options.](#)

Showing 1 - 20 of 144 code results for `filename:t_dm.R` in group

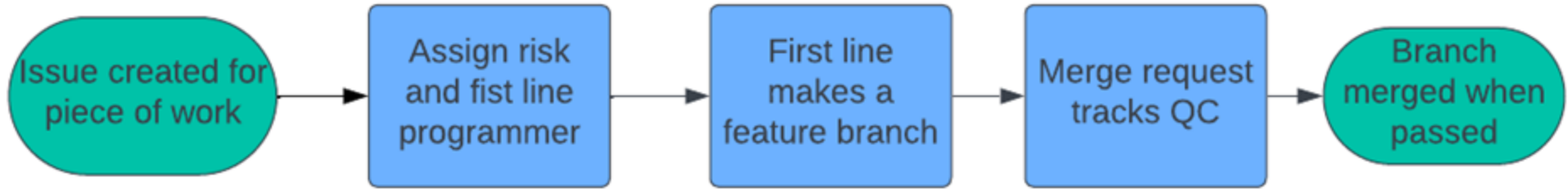


Project Management

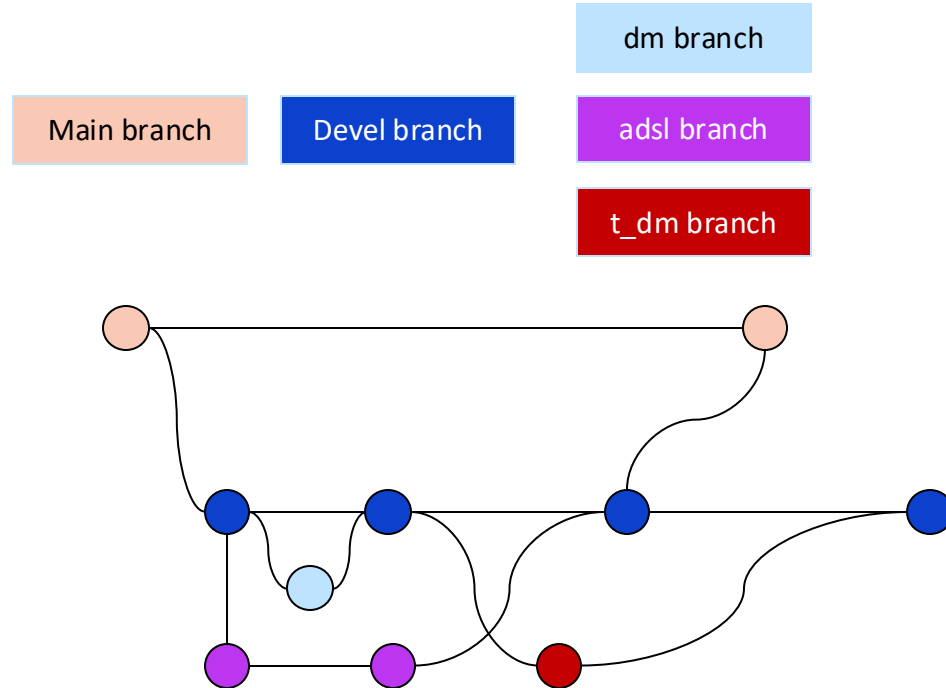


Our Git Strategy

Flow of work



Git Strategy



Challenges

Git is complicated



Gitlab assumptions

Assumptions:



Software Engineering	Clinical Reporting QC
■ CI/CD checks code before merged	QC outputs , not code
■ Code Review and Unit Testing	Code Review less common than double programming

The Future



Change

Questions?



Doing now what patients need next