National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Best Practices in Software and Data

Sadie Bartholomew, Nicola Farmer, Laurents Marker, Weronika Osmólska

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

1. Communities and standards

Sadie Bartholomew

Computational Scientist, NCAS-CMS
(CMS = Computational Modelling Services)

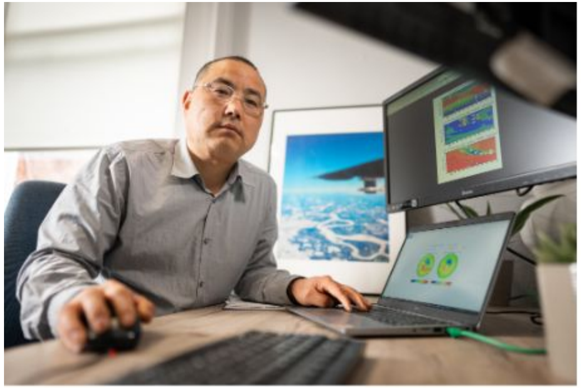
Software & data: vital to atmospheric & aligned science

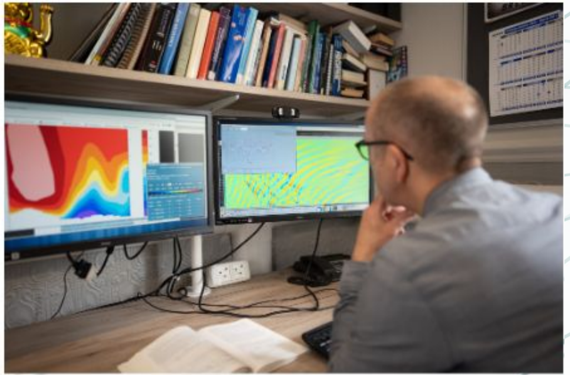
- Direct research (science, maths, etc.)
- Data analysis and visualisation
- Numerical modelling
- Atmospheric observations and measurements, e.g. ground-based lab or mobile instrumentation work, airborne measurements
- Data storage and archiving
- Computing (including high-performance - HPC) and workflows
- Infrastructure, services, user support
- And more…

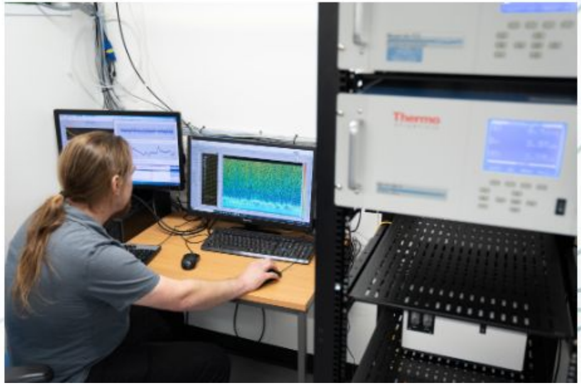
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

…so naturally are used widely & variously across NCAS

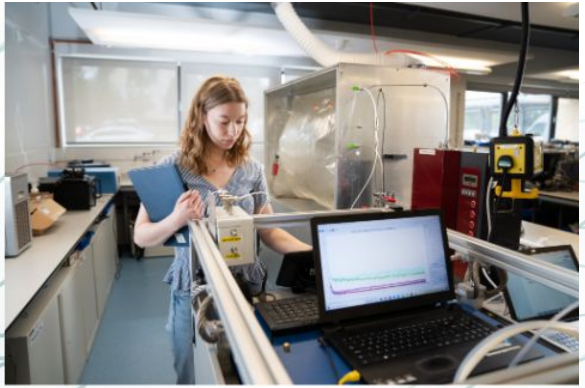
The photos that follow (and the previous one) are a selection taken from a NCAS Flickr album of various NCAS staff members conducting some of their typical work, where I have picked out photos which show software (data) in use

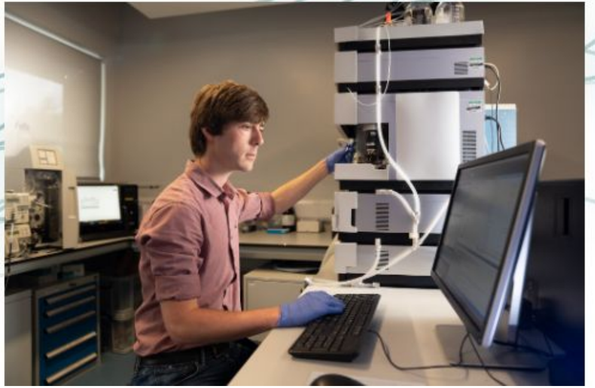






MIND YOUR HEAD





It helps to work together, not be (& remain) isolated*!

Notably, above all, we want to highly recommend:

1. **Finding relevant communities** to share results and challenges with, to find support from, etc.

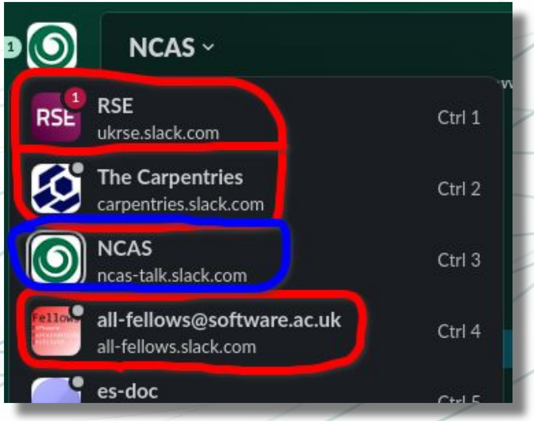
2. **Using standards (from the start!)**, so that we can work in a consistent and coherent way and maximise our ability to share code and data

*We are not containers (see Nicola's talk)

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Communities: my recommendations → general

1. For software (and data) in research, overall:
   - **Society of Research Software Engineering (society-rse.org)**: national society (+ community) with small membership fee per year but you can join the Slack channel, newsletter, events and conferences, etc. without being a member. Annual conference (RSECon) is brilliant!
   - **Software Sustainability Institute (www.software.ac.uk)**: "*dedicated to improving software in research*", with a newsletter; site including blog full of advice; workshops and events; fellowship programme.
   - **The Turing Way (book.the-turing-way.org)**: "*handbook to reproducible, ethical and collaborative data science*"
2. For training relating to software and data:
   - **The Carpentries (carpentries.org)**: "*teaches foundational coding and data science skills to researchers worldwide*". open workshops and lessons with online resources; discussions and more. Train to be an official instructor!



NCAS

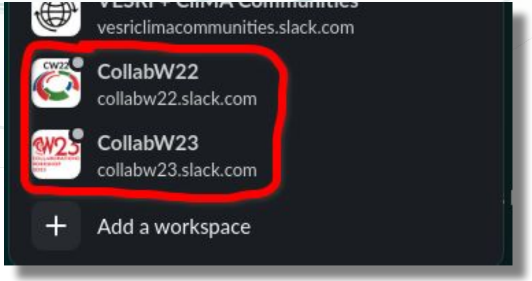
RSE
ukrse.slack.com
Ctrl 1

The Carpentries
carpentries.slack.com
Ctrl 2

NCAS
ncas-talk.slack.com
Ctrl 3

all-fellows@software.ac.uk
all-fellows.slack.com
Ctrl 4

es-doc
Ctrl 5



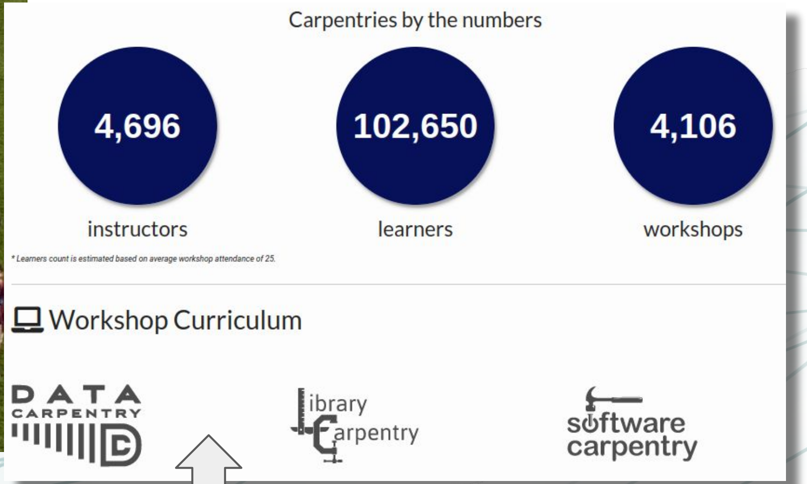
VESRI + CliMA Communities
vesriclimacommunities.slack.com

CollabW22
collabw22.slack.com

CollabW23
collabw23.slack.com

+ Add a workspace

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL



Carpentries by the numbers

4,696	102,650	4,106
instructors	learners	workshops

* Learners count is estimated based on average workshop attendance of 25.

Workshop Curriculum

DATA CARPENTRY

library carpentry

software carpentry

BETTER SOFTWARE BETTER RESEARCH

THE CARPENTRIES

SOFTWARE SUSTAINABILITY INSTITUTE

The TURING WAY

National Centre for Atmospheric Science

NATURAL ENVIRONMENT RESEARCH COUNCIL

Communities: my recommendations → domain-specific

For software and/or data within earth and aligned science/research, more specifically:

- **NCAS resources! (ncas.ac.uk/our-services/computer-modelling-and-data)**: as per Barbara's talk yesterday, often available to those outside of NCAS as well as NCAS staff
- **Pangeo (pangeo.io)**: "*cultivate an ecosystem in which the next generation of open-source analysis tools for ocean, atmosphere and climate science can be developed, distributed, and sustained*", with a supported 'Big Data' library ecosystem, blog, forums, and more
- **European Geosciences Union (EGU) and American Geophysical Union (AGU) Informatics (and similar) Divisions (www.egu.eu/essi/home, connect.agu.org/essi/home):** both unions hold annual General Assemblies where usually over ~20,000 people gather(!) and have divisions dedicated to (quoting from the AGU site linked) "*issues of data management and analysis, large-scale computational experimentation and modeling, and hardware and software infrastructure needs*". Membership needed to access most benefits, but could easily be worthwhile.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Standards: some examples that may apply to your work

1. **Code** standards:

   Language-specific style (e.g. PEP8 for Python); design patterns (for OO coding etc.); linting, continuous integration, standard packaging, standard documentation sections, etc.

2. **Data** standards:

   Data models; formats (e.g. netCDF); for storage/archiving, discovery and access; use of database & schema; API (application programming interface) design; use of metadata; vocabulary & semantics; etc.

3. ***Meta*data** standards:

   Domain-specific, e.g. CF Conventions (& COARDS which it extends) for netCDF data, CIM (Common Information Model) for climate models, etc.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

To close this sub-talk: your recommendations!

We’d love to hear further recommendations for communities and standards, if anyone has any to share. You can share with the group now, towards the end of the talk, or during discussion/networking time.

National Centre for
Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

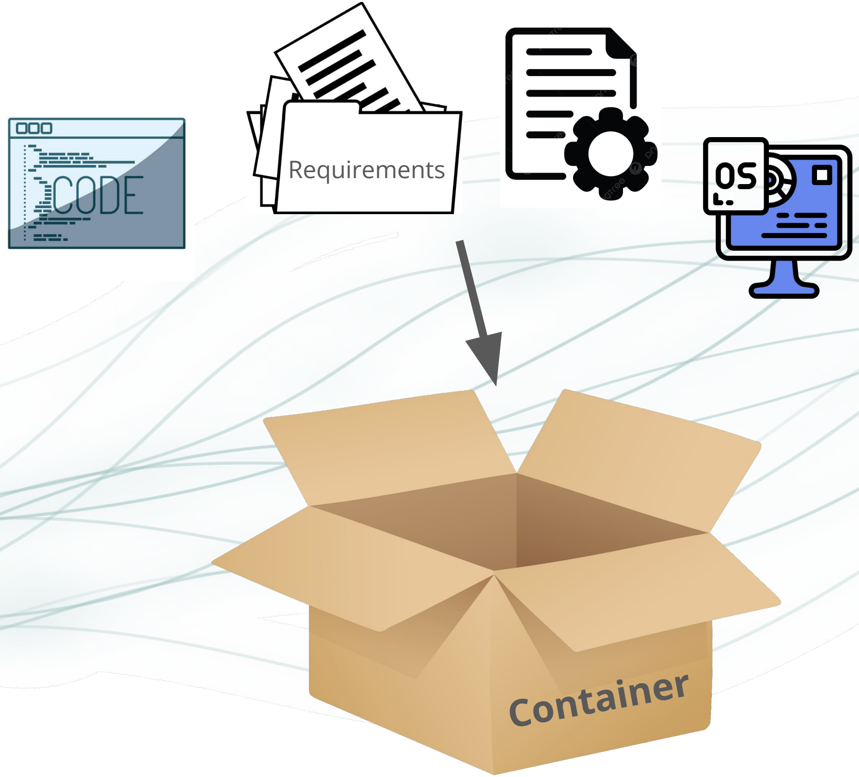
Containerisation

Nicola Farmer, Graduate Software Developer, CEDA

docker

What is containerisation?

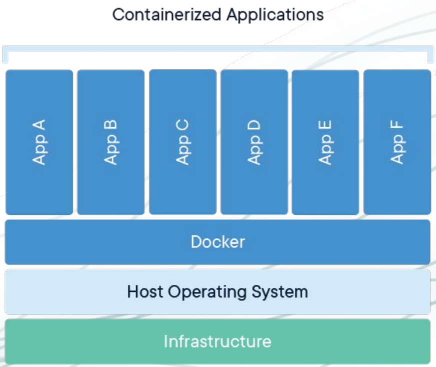
- It is essentially like packing up an application and everything it needs to run into a neat, portable box called a **container.**
- The container includes the application’s code, dependencies, settings and all the operating system components all bundled together.
- We use software called **Docker** to create containers, and software such as **Docker Compose** or **Kubernetes** to manage the containers.



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Containers vs Virtual Machines (VMs)

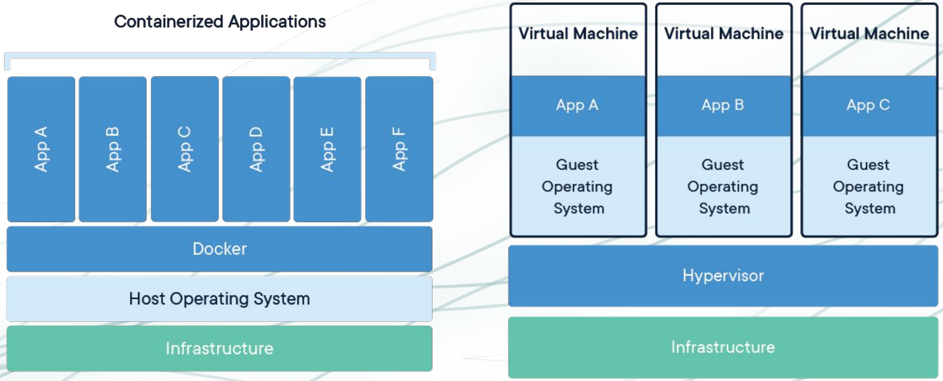
- Docker containers run on top of a **single host operating system (OS)** but have their own filesystem, network and namespace.



ⓘ A **virtual machine** simulates a real computer running its own operating system and applications

Containers vs Virtual Machines (VMs)

- Docker containers run on top of a **single host operating system (OS)** but have their own filesystem, network and namespace.
- VMs run on a **hypervisor layer** that abstracts the physical hardware and provides virtualized hardware to each VM - i.e. each VM has its own **guest OS.**



ⓘ A **virtual machine** simulates a real computer running its own operating system and applications

ⓘ A **hypervisor** manages and allocates resources on a physical computer to create and control multiple VMs

Why use containers?

Isolated

Isolated from the system they run on so no matter where you run the container (on your laptop, in a data-centre, in the cloud) it will run the same way without any conflicts or compatibility issues.

Lightweight

Containers are lightweight and fast to start up and shut down, making them great for deploying and scaling apps quick and efficiently. You can run multiple containers on the same host without sacrificing performance (depending on your architecture).

Portable

Containers are defined by a small text file so are more portable and resource-efficient than alternatives such as VMs. Allows apps to be written once and run anywhere.

How to create and use containers

Write your application

This could be any software project written in any language.

Create a Dockerfile

In your project directory create a file named Dockerfile. This contains the instructions for Docker on how to build your container image.

Use examples to build this!

Build a Docker image

Create a snapshot of your application and everything it needs to run.

Run your container

Tell your container to run the code inside it!

Access your application

Using the web browser or client depending on your app.

Docker is widely-used industry-standard software used to package up apps into containers.

An **Image** is a snapshot of your application and everything it needs to run.

For your first time, you may want to follow the tutorial on the Docker website that will guide you through it!

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Ways of using containers

docker Compose

Docker-compose built-in option to simplify the process of managing apps that consist of multiple Docker containers using a single configuration file **docker-compose.yml**. These containers work together to form a complete app with each container serving a different function (e.g. web server, database, cache). Useful for smaller apps, single-host deployments or local testing and development.

kubernetes

Kubernetes helps you manage, deploy and scale your containerised apps. Tell it what you would like to run and it handles the rest - e.g. automatic scaling, rolling updates and load balancing. Useful for larger apps, complex large-scale deployments across multiple servers.

CI-Pipeline: A series of automated steps you can set up to build, test and deploy and application. Can use tools like Github Actions or Gitlab. Can use container images in the pipeline to provide a consistent and reproducible environment.

Example Workflow: JASMIN accounts and projects portal

- Write the applications - React and Django apps
- Write a Dockerfile for each app
- Run each app (independently) using `docker run`
- Write a docker-compose.yaml file
- Run the apps together using `docker compose`

… and finally

- Write the files to scale up using Kubernetes

Resources and tips

- Docker: https://www.docker.com/
  - Software used to create containers
- Kubernetes: https://kubernetes.io/
  - Software used to manage containers and deploy app
- Helm: https://helm.sh/
  - Language used to manage Kubernetes applications
- Gitlab: https://about.gitlab.com/
  - Web-based platform supporting entire software dev lifecycle
- Github Actions: https://docs.github.com/en/actions
  - Helps automate software dev workflows within GitHub

- You can install Docker locally and play around with it using Docker desktop. Aim to just copy other Dockerfiles and get your app running in a container.
- Chat GPT can be quite useful for asking specific questions – it can help you write the commands you need and tell you what things mean!
- It makes more sense once you've gotten started and used it!
- The best way of learning is by example – find people working on the same thing and chat to them about it!

Questions

- Have any of you used these technologies before or something similar?
- How would you use containers in your work?
- Do you want to be supported with this and do you know where to get support?
- How can these resources be more accessible to you as part of NCAS?
- Any other questions?

Accessible research

I in 200 females and 2 in 25 males are affected by colour-vision deficiency [1].

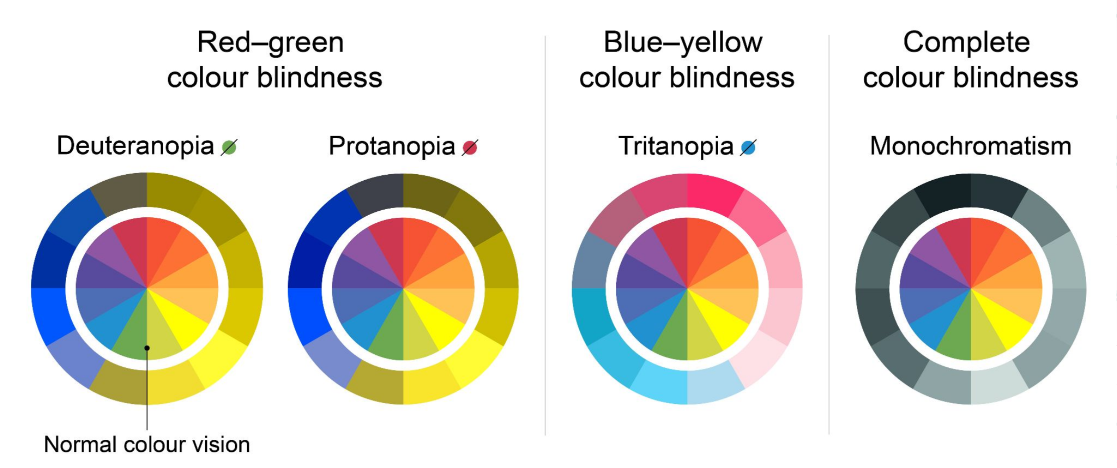
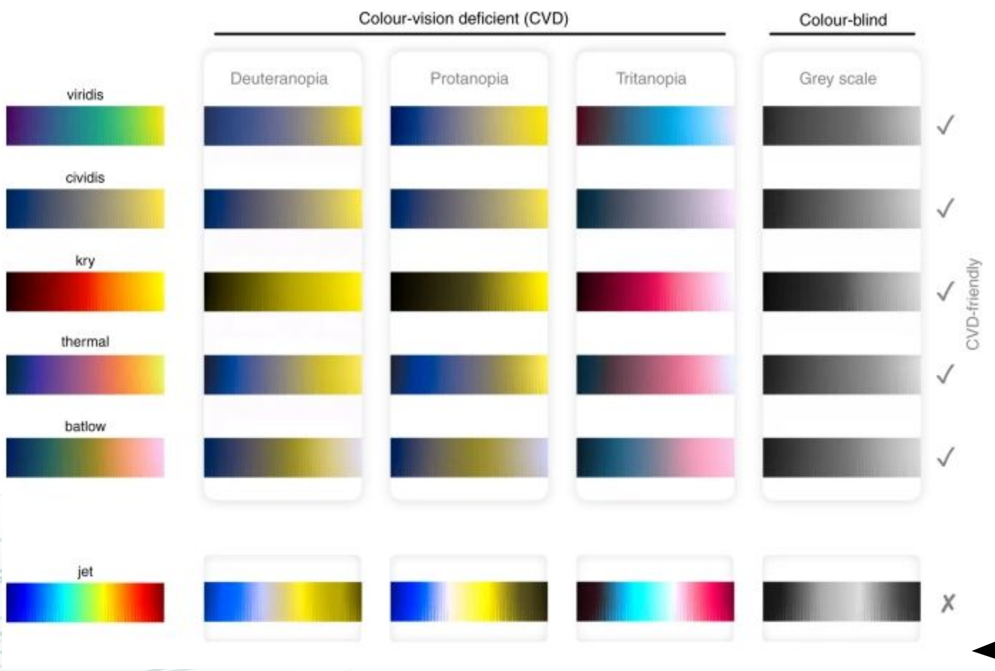
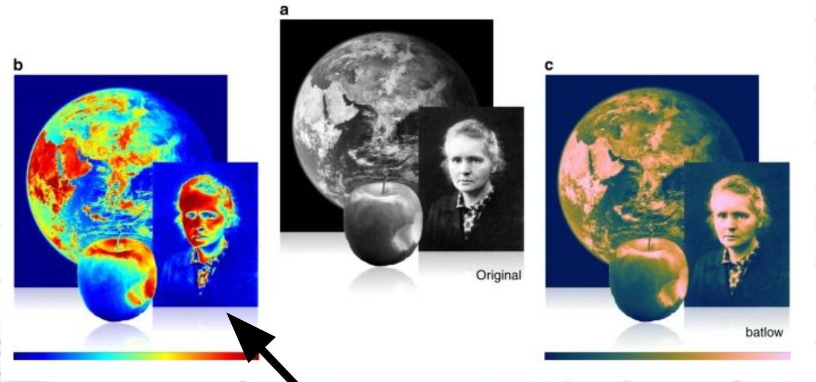

Red–green colour blindness
Blue–yellow colour blindness
Complete colour blindness
Deuteranopia
Protanopia
Tritanopia
Monochromatism
Normal colour vision

Figure credit health direct (https://www.healthdirect.gov.au/colour-blindness)

[1] ‘The Misuse of color in science communication’. Nature 2020

Accessible research





Distortion created by unscientific color maps (non uniform gradient color map).

Not color-vision friendly, but was the standard colormap of many softwares

Figures from: ‘The Misuse of color in science communication’. Nature 2020

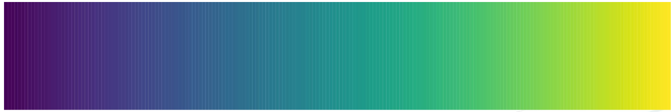
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Accessible research

Colorblind friendly colormaps:

(Python) Matplotlib:

-Viridis



-Cividis - looks identical to colorblind and non-colorblind individuals

Avoid:

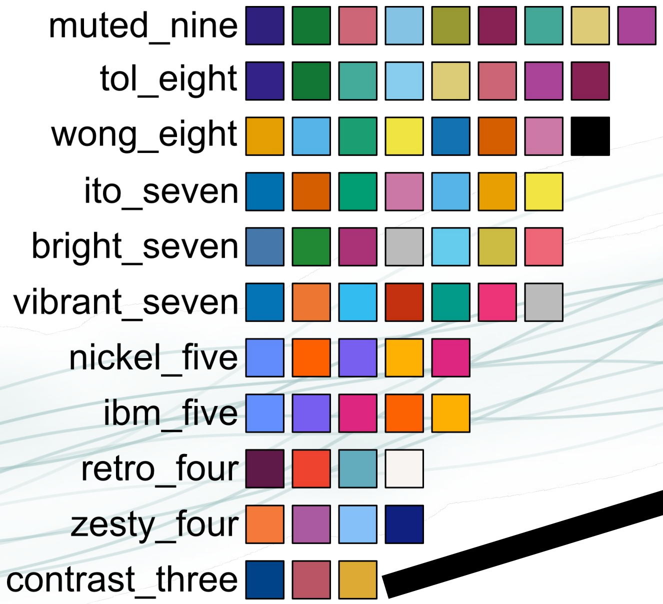
-Rainbow, jet colour palette

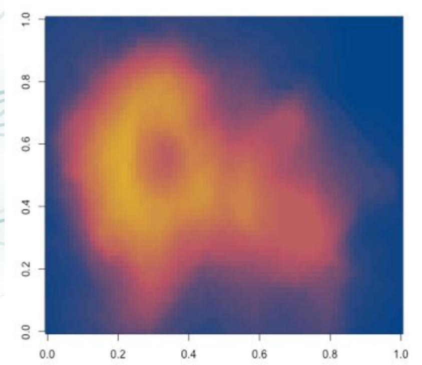
-Any green-red palette

-Any uneven color gradients

Accessible research

(R package) ggpubfigs (designed to be colorblind-friendly color palette):

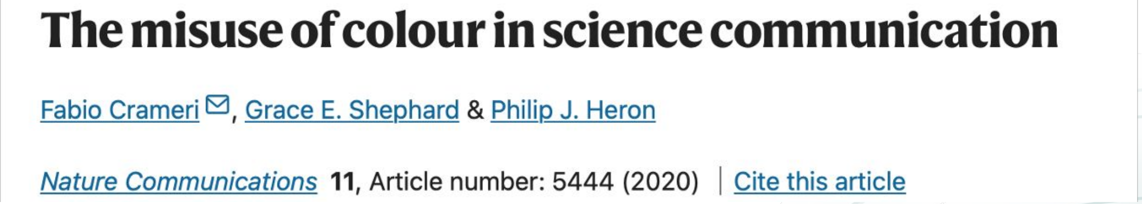




doi.org/10.1128/mra.00871-21 https://github.com/JLSteenwyk/ggpubfigs

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

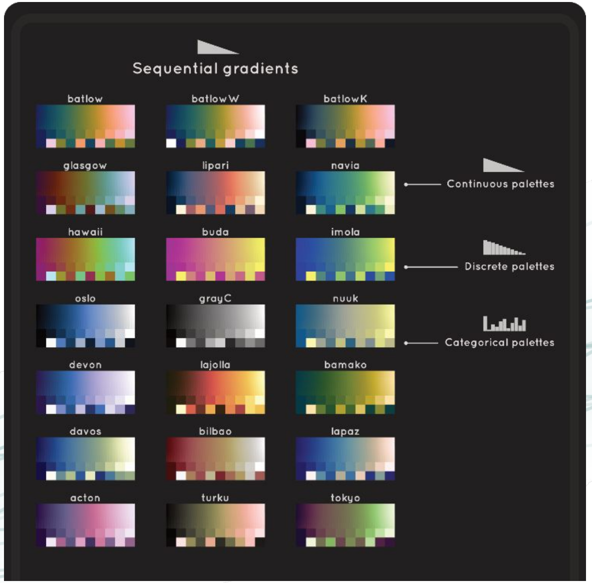
(Python, Matlab, Julia, R, GMT, QGIS, Ncview, Ferret, Plotly, Paraview, Visit, Mathematic, Gnuplot, Surfer, d3, SKUA-GOCAD, Petrel, XMapTools, Inskscape, Adobe photoshop…)

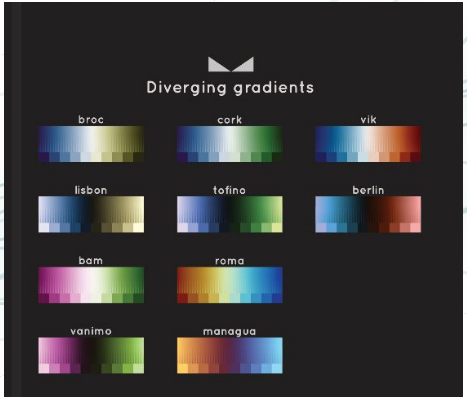


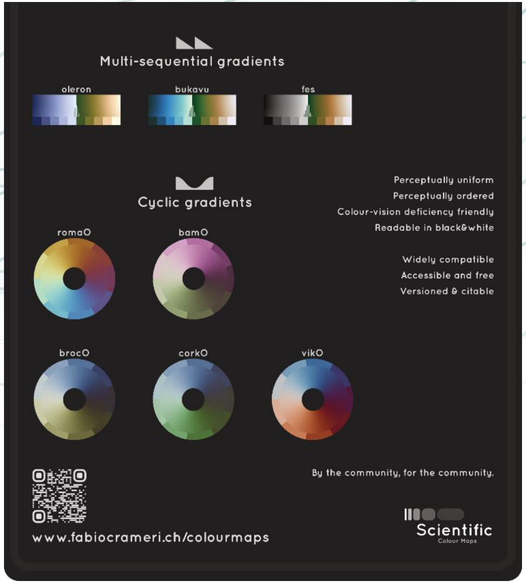
The misuse of colour in science communication

Fabio Crameri ✉, Grace E. Shephard & Philip J. Heron

Nature Communications **11**, Article number: 5444 (2020) | Cite this article







https://www.fabiocrameri.ch/colourmaps/

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

git and GitHub

Laurents Marker

Who here has used git before?

Who here uses git more than once a week?

Who here has used github before?

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Who here uses github more than once a week?

National Centre for
Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

What is git?

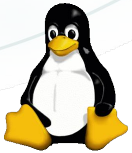
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

A version control system

> “Version control is a system that records changes to a file or set of files over time so that you can recall specific versions later.”

Pro Git Book, Scott Chacon and Ben Straub

Created by Linus Torvalds to aid in the development of the linux kernel in 2005



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

What does it do?

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

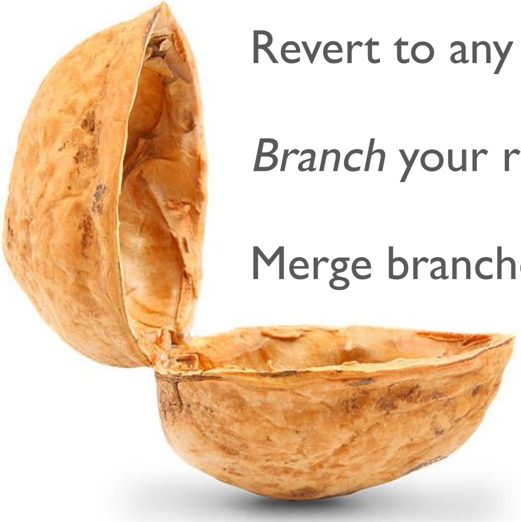
Organise your code for a project into a repository or *repo* (directory)

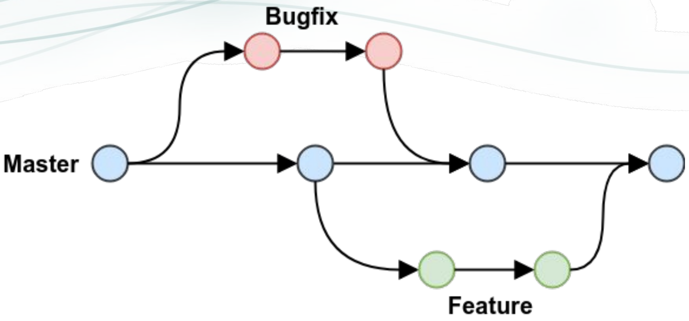
Commit changes that you make to your repository at some checkpoint

Revert to any of your previous commits

Branch your repository to develop new features or diverge code

Merge branches together to incorporate changes made in each





National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

What is github?

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

A version control system

"GitHub is a website and cloud-based service that helps developers store and manage their code, as well as track and control changes to their code."

What Is GitHub? A Beginner's Introduction to GitHub, kinsta.com

What does it do?

Stores and tracks git repositories on the cloud

Enables multiple users to have access to and edit repositories collaboratively by linking local copies to cloud instance

Pull requests to ensure code managers approve of changes to repositories

Lots of features to help developers!

Some useful stuff…

Submodules

A repo within a repo - repoception!

Link repo A into repo B as a directory within it

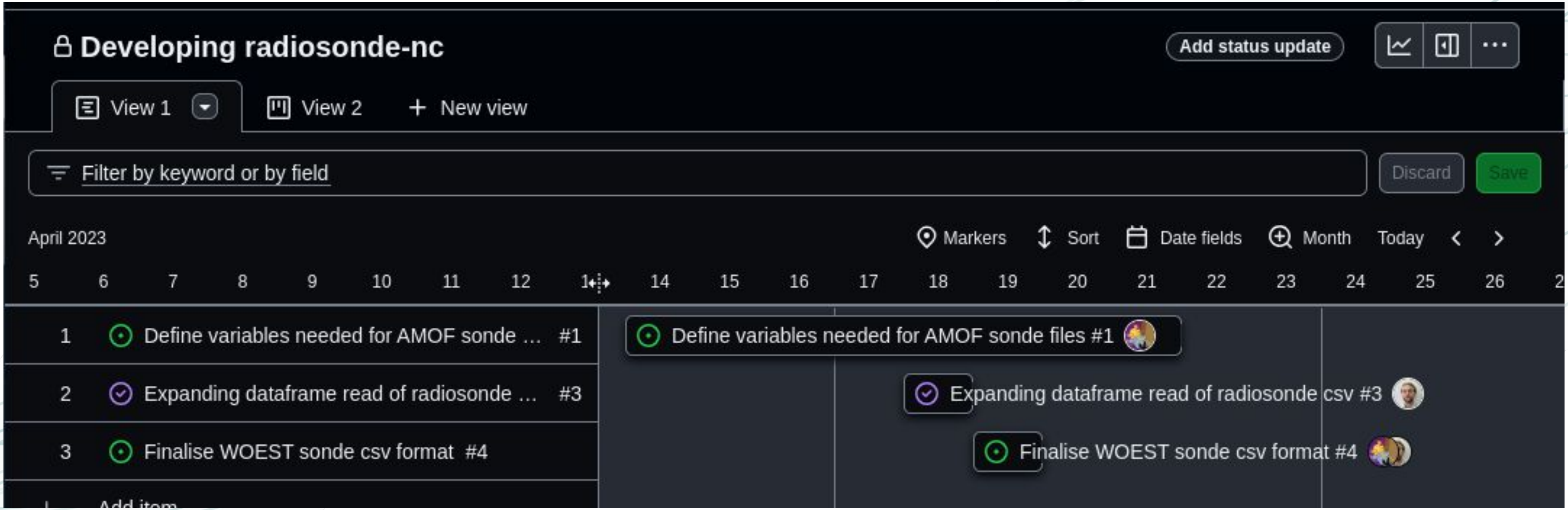
Very useful for breaking up and managing a big code base

Can manage inner repo separately and use for other projects

Good workflow for building robust utilities, packaging them with a repo and pulling them into bigger ones as submodules

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Projects and issues



1 Define variables needed for AMOF sonde … #1

2 Expanding dataframe read of radiosonde … #3

3 Finalise WOEST sonde csv format #4

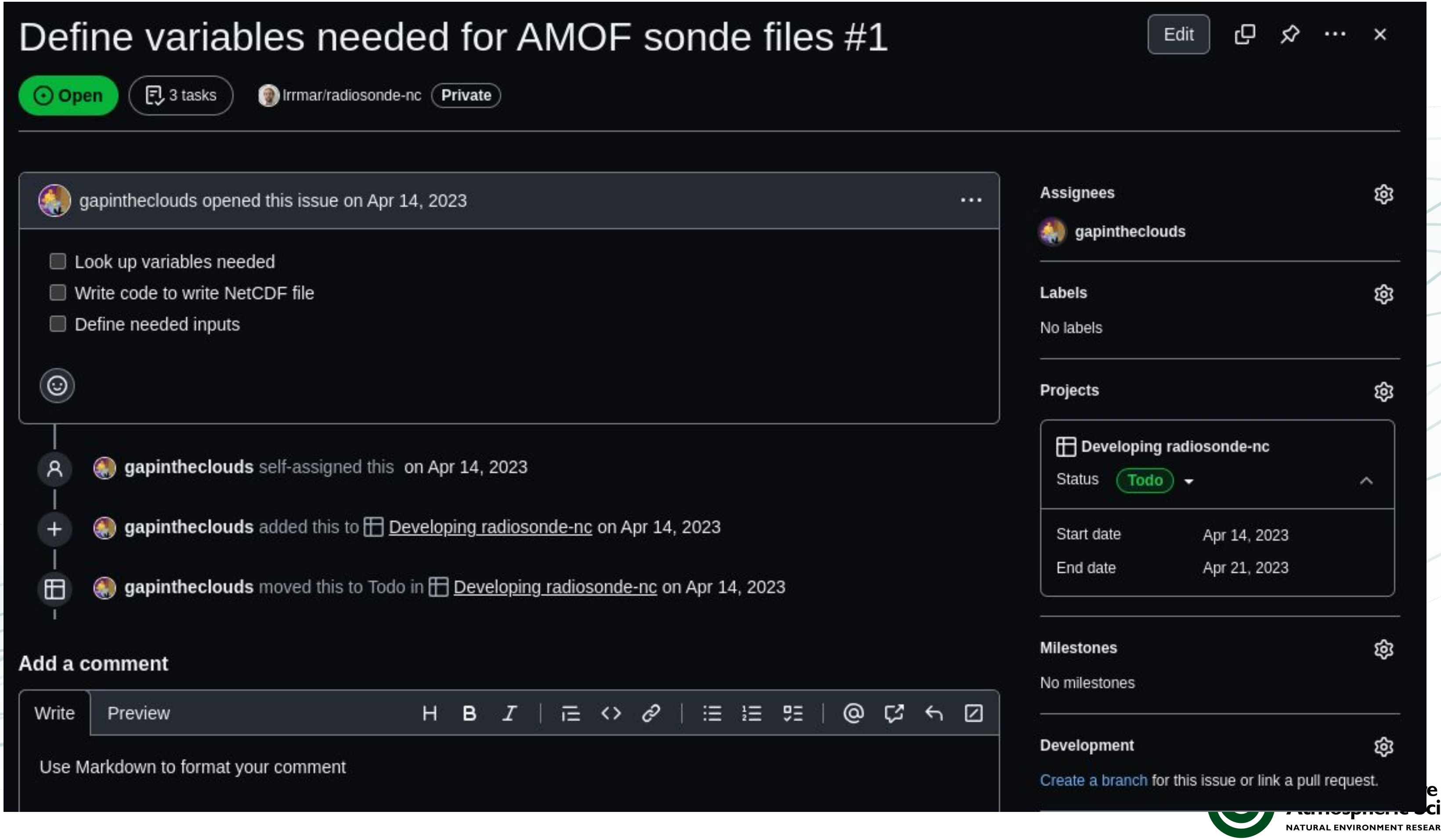
Add item

Define variables needed for AMOF sonde files #1

Expanding dataframe read of radiosonde csv #3

Finalise WOEST sonde csv format #4

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL



Define variables needed for AMOF sonde files #1

lrrmar/radiosonde-nc Private

gapintheclouds opened this issue on Apr 14, 2023

□ Look up variables needed

□ Write code to write NetCDF file

□ Define needed inputs

gapintheclouds self-assigned this on Apr 14, 2023

gapintheclouds added this to Developing radiosonde-nc on Apr 14, 2023

gapintheclouds moved this to Todo in Developing radiosonde-nc on Apr 14, 2023

Add a comment

Use Markdown to format your comment

Assignees

gapintheclouds

Labels

No labels

Projects

Developing radiosonde-nc

Status: Todo

Start date: Apr 14, 2023

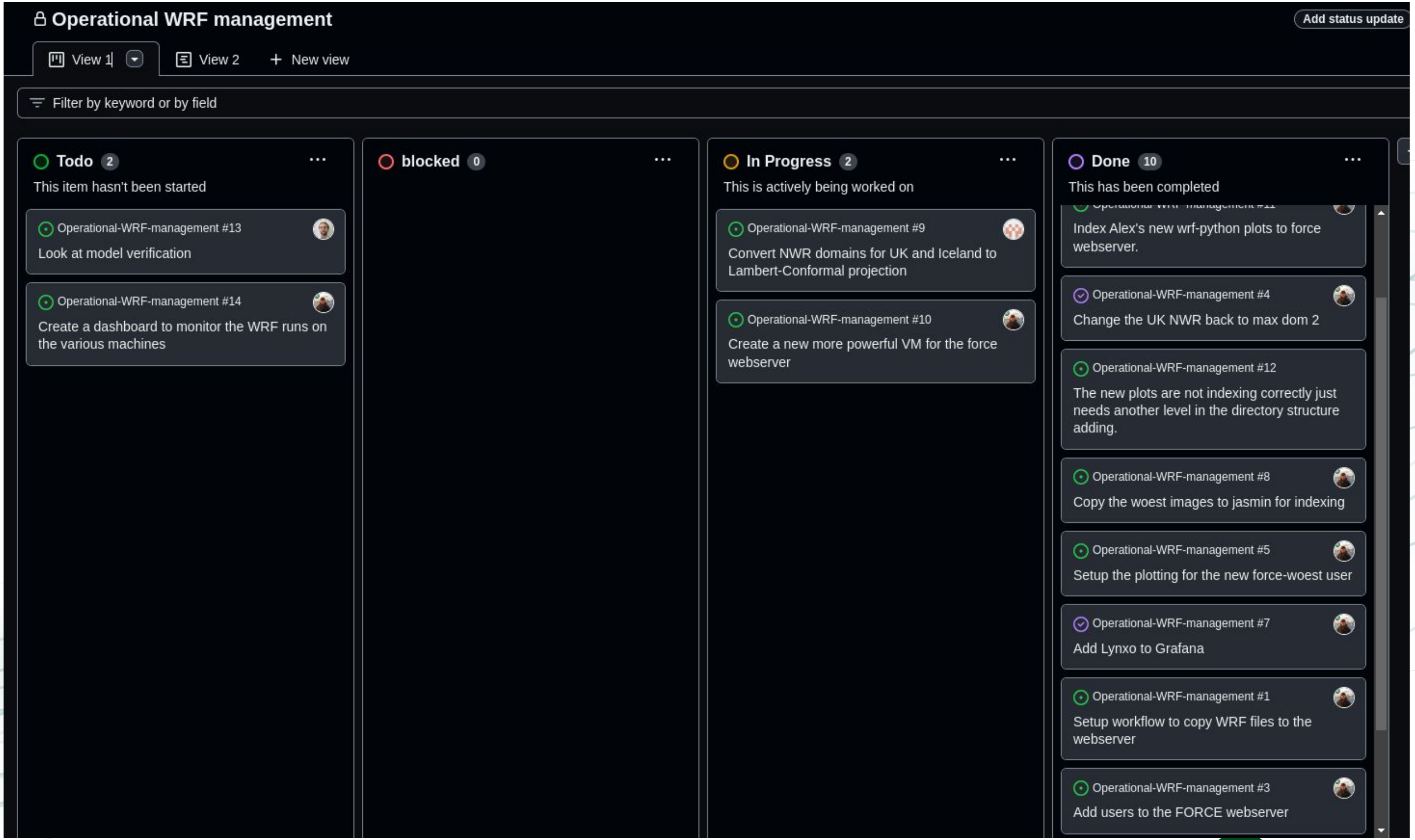
End date: Apr 21, 2023

Milestones

No milestones

Development

Create a branch for this issue or link a pull request.



Todo 2

This item hasn't been started

- Operational-WRF-management #13
  Look at model verification
- Operational-WRF-management #14
  Create a dashboard to monitor the WRF runs on the various machines

blocked 0

In Progress 2

This is actively being worked on

- Operational-WRF-management #9
  Convert NWR domains for UK and Iceland to Lambert-Conformal projection
- Operational-WRF-management #10
  Create a new more powerful VM for the force webserver

Done 10

This has been completed

- Operational-WRF-management #11
  Index Alex's new wrf-python plots to force webserver.
- Operational-WRF-management #4
  Change the UK NWR back to max dom 2
- Operational-WRF-management #12
  The new plots are not indexing correctly just needs another level in the directory structure adding.
- Operational-WRF-management #8
  Copy the woest images to jasmin for indexing
- Operational-WRF-management #5
  Setup the plotting for the new force-woest user
- Operational-WRF-management #7
  Add Lynxo to Grafana
- Operational-WRF-management #1
  Setup workflow to copy WRF files to the webserver
- Operational-WRF-management #3
  Add users to the FORCE webserver

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL