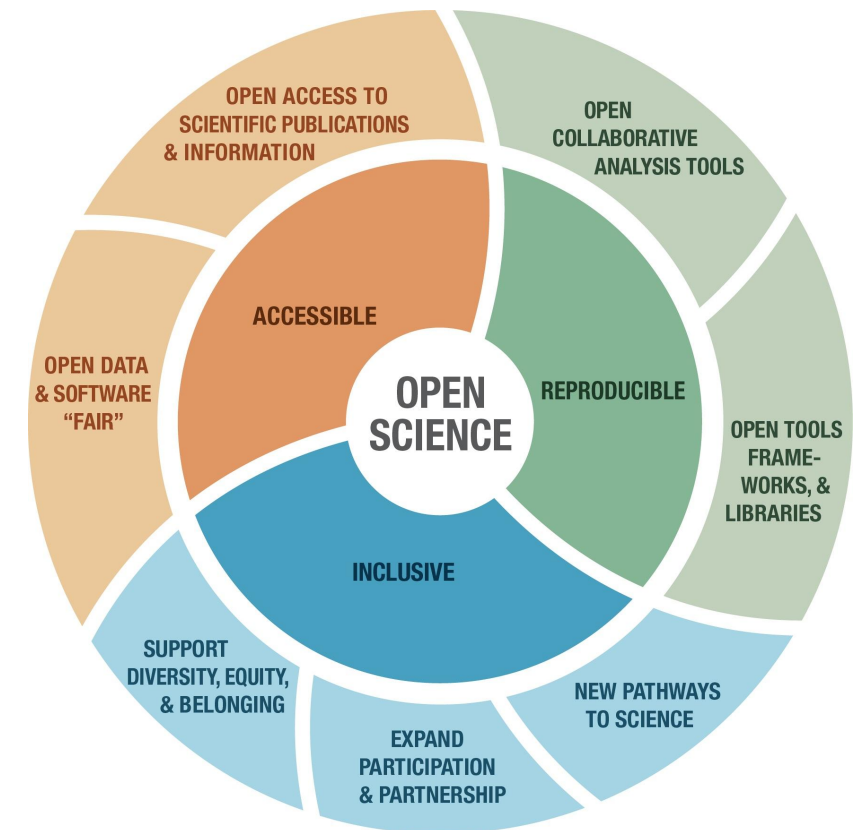


Community weather radar software and Open Science

Daniel Michelson and Max Grover
European Conference on Radar in
Meteorology and Hydrology
August 28, 2022
Locarno, Switzerland



Outline

- Open Science basics and background (bottom-up)
- Open Science developments
- Moving towards more Open Science practices

Disclaimer: this talk is not intended as communication of direction on how science should be conducted, nor is it intended as critique of existing practices.

Not **what** to science, but ideas on **how** to science.

What is Open Science in practice?

Open Data + Open Algorithms + Open Software + Open Peer Review + Open Access Publication

Thereby, science process and outputs can be:

- ✓ Transparent
- ✓ Reproducible
- ✓ Transferrable
- ✓ Collaborative
- ✓ Credible!
- ✓ Durable, sustainable



IMPORTANT!

No conflict between Intellectual Property Rights (IPR) and licensing.

NASA Transform to Open Science (TOPS) Initiative

- Key components
 - Open meetings
 - Open curriculum
 - Open data
 - Open compute
 - Incentives for participating
- Goals
 - Increase understanding and adoption of open science principles and techniques in our Mission and Research Communities
 - Accelerate major scientific discoveries through supporting the adoption of open science
 - Broaden participation by historically excluded communities
- \$40 million+ initiative

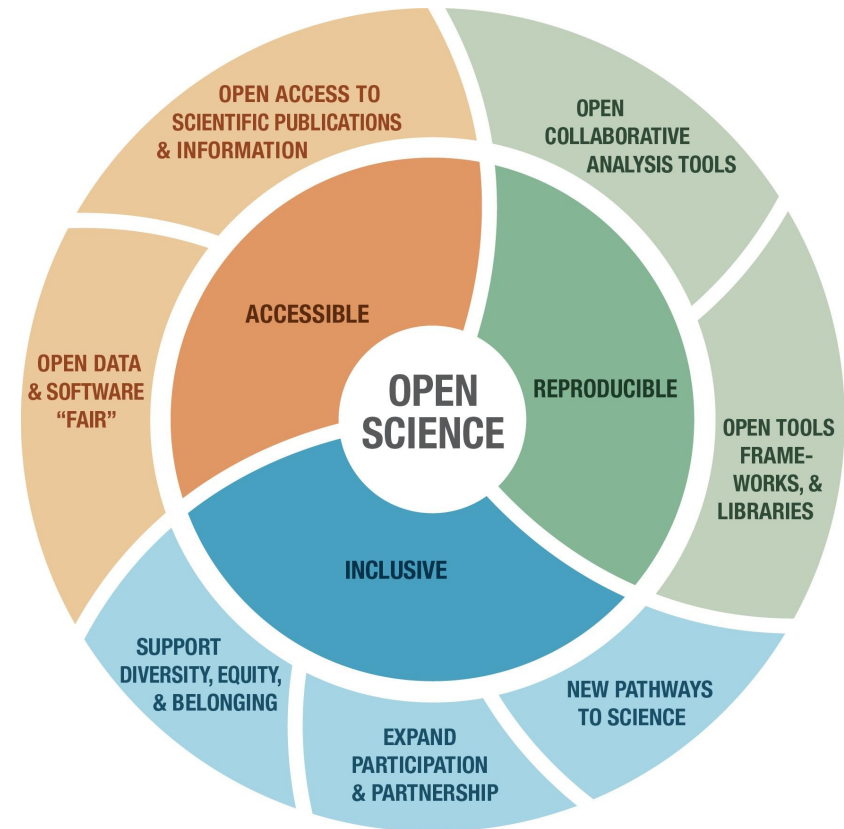


Image courtesy of NASA TOPS

What is Pangeo?

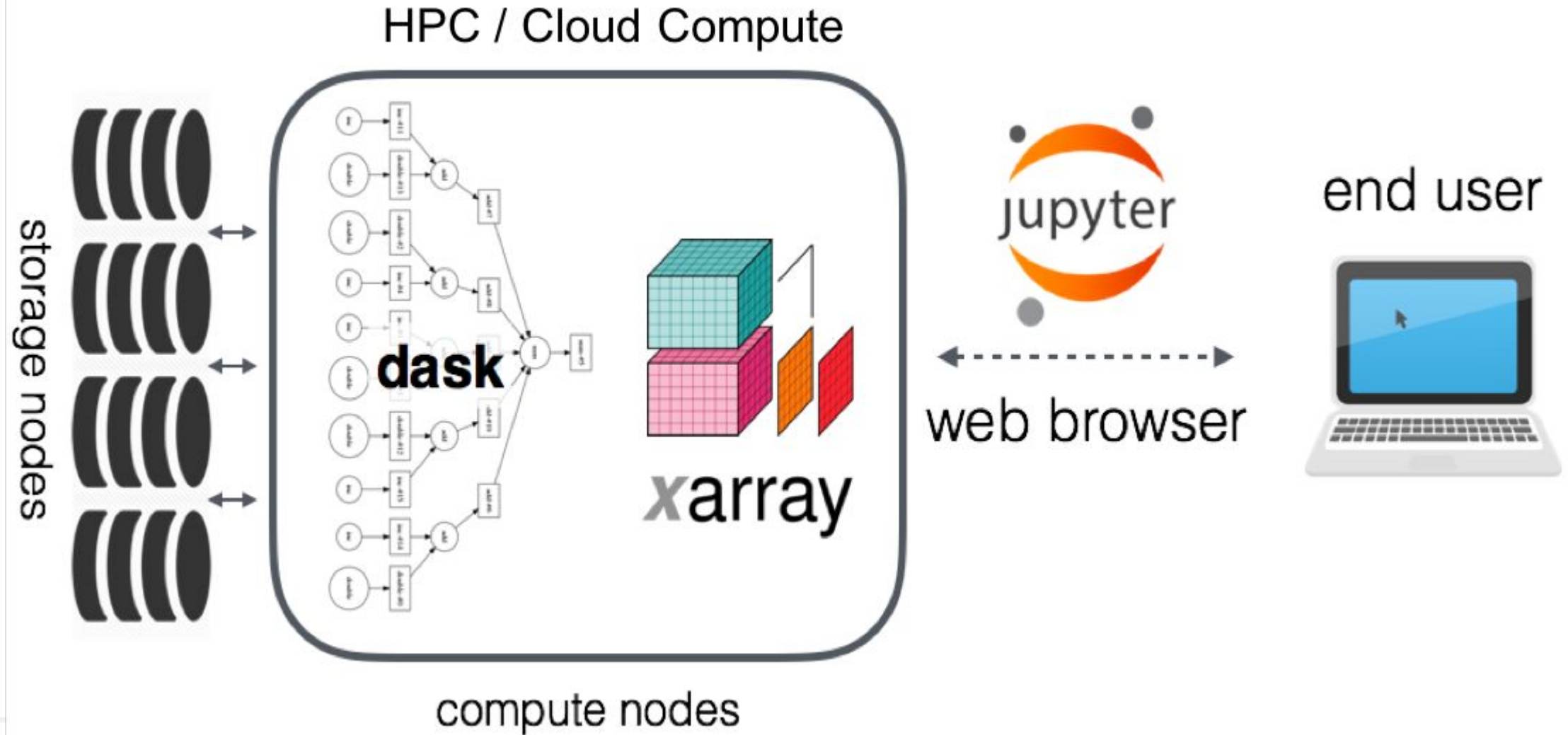


PANGEO

A community of scientists and software engineers with the goal to:

- Foster **collaboration** around the **open source**, scientific Python ecosystem.
- **Support the development** of domain-specific geo-science packages.
- Improve **scalability** of these tools to handle terabyte-scale datasets on **cloud** and **HPC** platforms.

Pangeo







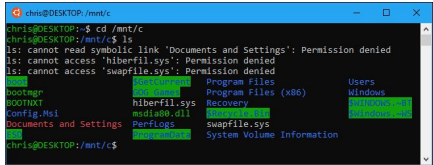





The Importance of Interoperability

INTERCHANGEABLE PIECES IN PANGEO (PICK 1 OR MORE FROM EACH ROW)

Data Models	 xarray	 Iris	 pandas $y_i = \beta^T x_i + \mu_0 + \epsilon_i$
N-D Arrays	 NumPy	 DASK	
Processing Mode	Interactive  jupyter	Batch 	Serverless 
Compute Platform	HPC 	 aws	 Google Cloud Platform
Foundation	 python™		

Applying this to the Open Radar Community

(Pick one or more from each row)

Data Models	PyART Radar (based on CfRadial 1.4) BALTRAD objects (based on ODIM_H5)	wradlib 
N-D Arrays		
Processing Mode	Interactive 	Batch 
Compute Platform	HPC 	Cloud 
Foundation + Language	 python™ Py-ART, wradlib, PyRAD	  LROSE, BALTRAD

Who is a scientist?



Dr Esther
@EstOdek



QUESTION: What makes someone a scientist?

I'll start with one:

If you 'do' the science (i.e. generate the data) then you're a scientist.

Period. It doesn't matter if you are a bachelors/masters/PhD student. If you DO the science, you are a scientist.

13:11 · 2019-05-02 · [Twitter for iPhone](#)

- Affects incentive structures: who can and should receive credit for publishing not only papers but also data sets and code.
- Career progression, opportunity



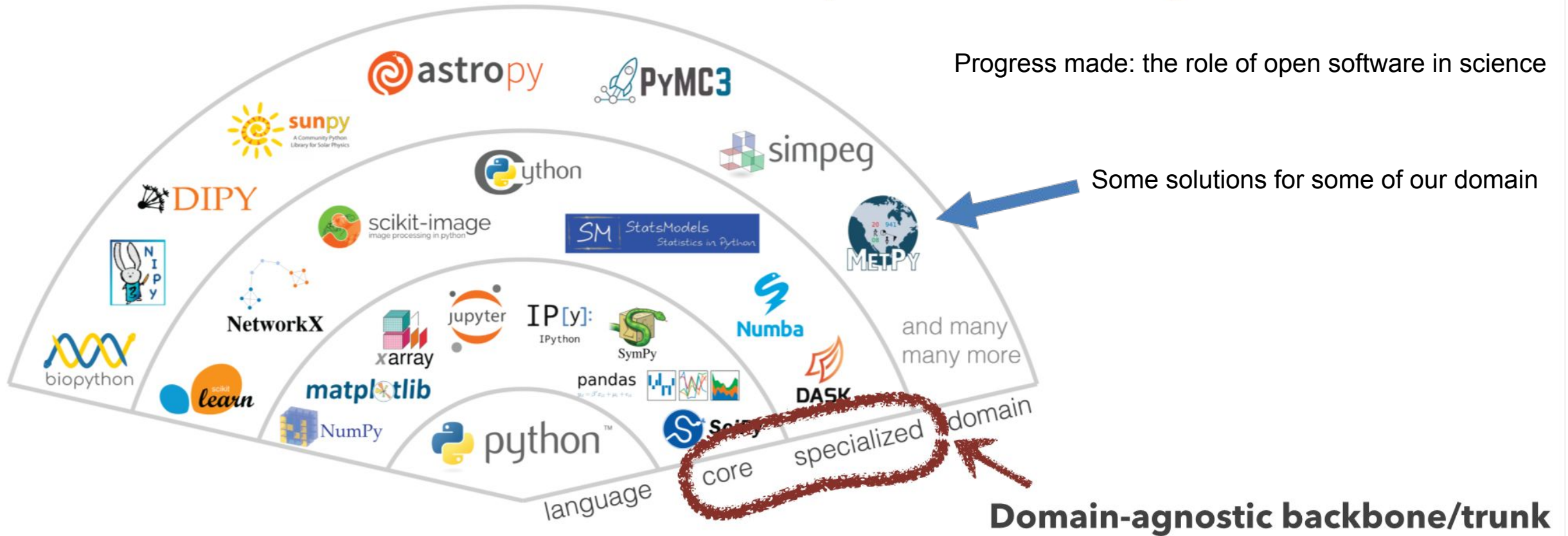
Ivan Alcantara
@ivancedric_a



I joined twitter to share and read scientific papers and now I'm an anarchist

13:50 · 2020-06-12 · [Twitter for iPhone](#)

"The Stack": a complete ecosystem



Lindsey Heagy
Fernando Pérez
UC Berkley

Domain-agnostic backbone/trunk

- Not “real CS”
- Not “real research”
- Nobody’s problem
- Yet critical to everybody else

Open science solutions – pro tips (1/2)

1. Use Git for collaborative change/version code management
 - GitHub: github.com
 - GitLab: gitlab.com
2. Linux virtual machine on your local computer
 - Cost-effective near-replication of official environments
 - Good for development, even while traveling
 - VM is the starting point, a vehicle, for creating transferrable computational environments
 - Today's cloud instance is an “elaborated” VM

Open science solutions – pro tips (2/2)

3. If you have a choice, use an open programming language
4. Avoid proprietary algorithms like Numerical Recipes
5. Publish code
6. Publish data
7. Publish code & data associated with a study/paper
How? (next slides)
7. Publish in Open Access journals



Eric Green
@ericpgreen

In response to a request to see the data and code during peer review, the authors replied that the request was highly unusual and never encountered in hundreds of articles they published. Here's to making more progress in 2019. opennessinitiative.org
[#openscience](https://twitter.com/ericpgreen)

08:52 · 2018-12-30 from [Durham, NC](#) · [Twitter for iPhone](#)

Publishing code and data for a study/paper (1/2)

1. Put code, data and ancillary scripts on GitHub
2. “Create a release” in GitHub

Branch: master

Go to file Add file Code

Daniel Michelson committed cd7df2f on 30 Dec 2019 5 commits 1 branch 1 tag

EMET_plots	Adding content	6 months ago
EMET_verif	Adding content	6 months ago
radar_plots	Adding content	6 months ago
rave_ec	ADD: rave_ec build script	6 months ago
LICENSE	Adding content	6 months ago
README.md	Update README.md	6 months ago

README.md

drqc_article

Daniel Michelson
Observation Based Research Section
Meteorological Research Division
Atmospheric Science and Technology Directorate
Science and Technology Branch
Environment and Climate Change Canada

This project contains scripts, routines, and references to projects used to generate the results documented in the DRQC article.

The weather radar data processing system used is the BALTRAD Toolbox together with several of its associated software packages and ECCC add-ons. This system is built into an operationally-deployable so-called SSM package. The SSM is built using the build_ubuntu14.04_ssm_EC.sh script which downloads, configures, builds, tests, and installs the software. This is done in most cases by checking out explicit versions of the packages in question, since all packages are managed either from the publically accessible BALTRAD Git server, GitHub, or a GitLab instance managed internally by ECCC.

A snapshot of the internal rave_ec project is included here for the purposes of creating an archive repository.

About

Scripts, routines, and references to projects used to generate the results documented in the DRQC article

Readme

View license

Releases 1

DRQC Article Latest on 30 Dec 2019

Packages

No packages published
[Publish your first package](#)

Languages

Python 81.9% Shell 6.2% C 5.8% PHP 3.5% Makefile 2.6%

Publishing code and data for a study/paper (2/2)

3. Create persistent Digital Object Identifier (DOI) of the release

zenodo

Search

Upload

Communities

Log in

Sign up

December 30, 2019

Software

Open Access

DanielMichelson/drqc_article: DRQC Article

Daniel Michelson

First version of the release of the repository created to store code and plotting routines used to create the article documenting DRQC and methods used to monitor the impact of quality control on the quality of weather radar data.

Preview

drqc_article-v1.0.zip

DanielMichelson-drqc_article-cd7df2f

EMET_plots

82_new_qc

82_new_qc_far.png 88.8 kB

82_new_qc_far.txt 9.7 kB

82_new_qc_far_31-1-426-20190711222425.png 87.5 kB

82_new_qc_fbi.txt 9.7 kB

82_new_qc_fbi_31-1-427-20190711222443.png 88.8 kB

82_new_qc_hss.txt 9.7 kB

82_new_qc_hss_31-1-428-20190711222501.png 90.3 kB

82_new_qc_pod.txt 9.7 kB

82_new_qc_pod_31-1-430-20190711222537.png 96.2 kB

82_new_qc_pss.txt 9.7 kB

82_new_qc_pss_31-1-431-20190711222555.png 93.2 kB

93_old_qc

93_old_qc_far.txt 8.3 kB

93_old_qc_far_23-1-784-20180720202649.png 90.4 kB

93_old_qc_fbi.txt 8.2 kB

93_old_qc_fbi_23-1-785-20180720202713.png 90.4 kB

Files (58.0 MB)

Name	Size
DanielMichelson/drqc_article-v1.0.zip	58.0 MB

md5:8ecb1f3d347716777a321286f04c79fd

Preview

Download

Beta Citations 0

Show only:

Literature (0)

Dataset (0)

Software (0)

Unknown (0)

Citations to this version

Search

No citations.

19 views

2 downloads

See more details...

Available in

GitHub

Indexed in

OpenAIRE

Publication date:

December 30, 2019

DOI:

DOI 10.5281/zenodo.3595509

Related identifiers:

Supplement to

https://github.com/DanielMichelson/drqc_article/tree/v1.0

License (for files):

Other (Open)

Versions

Version v1.0

10.5281/zenodo.3595509

Dec 30, 2019

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.3595508. This DOI represents all versions, and will always resolve to the latest one. Read more.

“A minimum standard for publishing...”

- Create a section in the paper entitled e.g. “Computation procedures”
- Add a statement describing code availability and accessibility
 - Refer to your DOI

are also only used because the differences in quality control using the DRQC only apply to radar data from the NEXRAD network.

2.3 | Computational procedures

The software environments used to generate the results in the present paper were collected into a project repository, publicly accessible at Michelson (2019). Part of this project is to preserve specific versions of the procedures used in an effort to achieve reproducibility according to Irving (2015).

The weather radar data-processing environment is the BALTRAD¹ Toolbox (Michelson *et al.*, 2018). This is

one of several open-source software environments that have emerged in recent years supporting the weather radar community (Heistermann *et al.*, 2014), even integrated together into working, publicly available solutions (Heistermann *et al.*, 2015).

Figures were created using Matplotlib (Hunter, 2007), which is the standard plotting environment in the Python programming language. Scientific colour maps used with plotting radar data were based on those provided by Crameri (2018), with the goal of displaying relative differences uniformly with their continuous scales. Additionally, Pandas (McKinney, 2017) was used to prepare the plots containing skill score results.

Damian Irving, 2016:

**A Minimum Standard for Publishing
Computational Results in the Weather and
Climate Sciences**

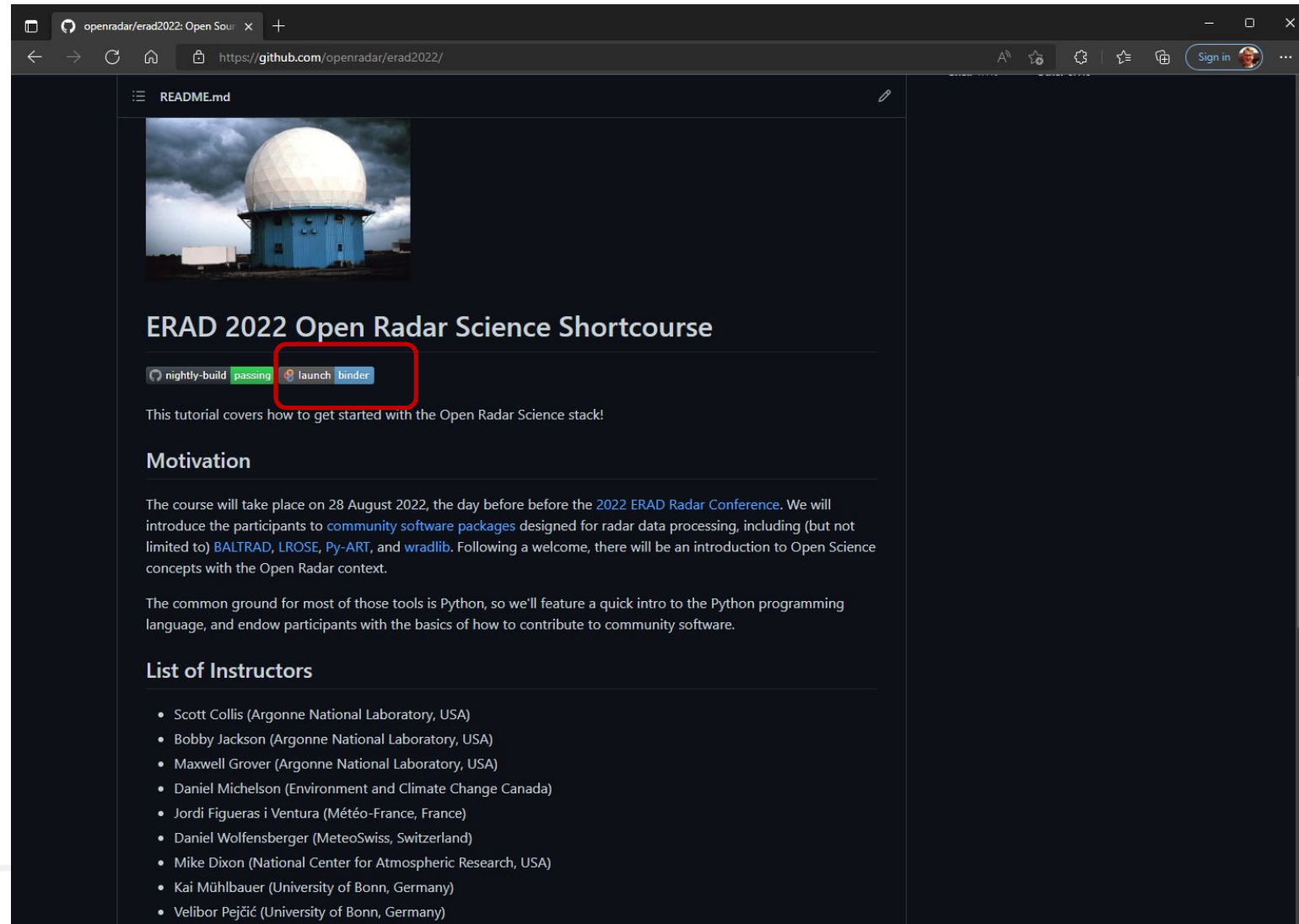
<https://doi.org/10.1175/BAMS-D-15-00010.1>

Michelson, D. (2019) DRQC Article. Zenodo. <http://doi.org/10.5281/zenodo.3595509>.

Advanced pro tips: open workflows (1/2)

- Same logic used to build local VM can be scaled up to HPC and cloud computing platforms
 - Pangeo cloud and mybinder.org are good examples: automated deployment!
 - Jupyter notebooks can be used to organize:
 - Educational material, proof-of-concept, little resources needed
 - Parallel cloud computing
 - Google Colab can run Jupyter notebooks found on GitHub.
 - If data are also in the cloud, e.g. NEXRAD LII radar data in AWS, then end-to-end workflows are achievable openly!
-

Advanced pro tips: open workflows (2/2)



The screenshot shows a web browser window displaying the GitHub repository page for 'openradar/erad2022: Open Sour'. The page title is 'ERAD 2022 Open Radar Science Shortcourse'. Below the title, there is a 'nightly-build' button with a 'passing' status, and a 'launch binder' button which is highlighted with a red rectangle. The page content includes a 'Motivation' section and a 'List of Instructors' section.

ERAD 2022 Open Radar Science Shortcourse

[nightly-build](#) passing [launch binder](#)

This tutorial covers how to get started with the Open Radar Science stack!

Motivation

The course will take place on 28 August 2022, the day before before the 2022 ERAD Radar Conference. We will introduce the participants to [community software packages](#) designed for radar data processing, including (but not limited to) [BALTRAD](#), [LROSE](#), [Py-ART](#), and [wradlib](#). Following a welcome, there will be an introduction to Open Science concepts with the Open Radar context.

The common ground for most of those tools is Python, so we'll feature a quick intro to the Python programming language, and endow participants with the basics of how to contribute to community software.

List of Instructors

- Scott Collis (Argonne National Laboratory, USA)
- Bobby Jackson (Argonne National Laboratory, USA)
- Maxwell Grover (Argonne National Laboratory, USA)
- Daniel Michelson (Environment and Climate Change Canada)
- Jordi Figueras i Ventura (Météo-France, France)
- Daniel Wolfensberger (MeteoSwiss, Switzerland)
- Mike Dixon (National Center for Atmospheric Research, USA)
- Kai Mühlbauer (University of Bonn, Germany)
- Velibor Pejčíc (University of Bonn, Germany)

Rethinking the roles of the scientific paper and scientific process

- The paper:
 - ✓ *advertises* the science
 - ✓ provides analysis and interpretation
 - ✓ peer review gives it the seal of approval
 - The *substance* of the science is in the data, code and workflow
 - ✓ Because everything is open, transparent, & reproducible
 - Working technical solutions can be collaborative from the start
 - Innovation cycles can accelerate (COVID-19 vaccines?)
 - People who do the heavy lifting on preparing and publishing data and code can be potentially rewarded for it
-

Thank you! Now Get Involved!

[Home](#)[Projects](#)[Community](#)[News](#)[Virtual Machine](#)[Open data](#)

The weather radar software community

A community platform for free and open source weather radar software.



What is this about?

- You don't want to pay license fees?
- You don't want to commit to a black box software?
- You are concerned about the PhD student that comes after you?
- You want to efficiently collaborate with people around the world?
- You want to stand on the shoulders of giants instead of reinventing the wheel?
- You don't care if it's free, you just want it to work?

If it's a **yes** to any of these questions, then have a closer look...

Social

[GitHub](#)[Gitter](#)

Categories

[News](#)[Projects](#)

Links

[ERAD2016](#)[ERAD2018](#)[ERAD2022](#)

openradarscience.org