

## Open Science: Yes You Can

Cédric H. David 2022-11-09 NASA Earth Surface and Interior Solid Earth Team Meeting, La Jolla, CA.



© 2022 Jet Propulsion Laboratory, California Institute of Technology

This document has been reviewed and determined not to contain export controlled technical data.

# The components of Open Science

"This AGU journal wants me to publish my code and data, I have no clue how"

"My team members keep building the same code to analyze the same data, I wish there was a better way"

"Can you believe how these authors analyzed data from my satellite? They have no clue!"

"Sorry, that was like 10 years ago, I don't remember this specific detail of my analysis"

"I'm afraid to update my code, it might break things"

"Can't do open science at NASA, I'll get fired!"



## Open science, you don't have to do it (or do you?), but you might learn some tricks regardless

	FROM:	EXECUTIVE OFFICE OF THE PRESIDENT OFFICE OF SCIENCE AND TECHNOLOGY POLICY WASHINGTON, D.C. 20002 February 22, 2013 UM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES John P. Holdren JULL Director JULL Increasing Access to the Results of Federally Funded Scientific Research
		EXECUTIVE OFFICE OF THE PRESIDENT OFFICE OF SCIENCE AND TECHNOLOGY POLICY WASHINGTON, D.C. 20502
1. Paper		August 25, 2022
	MEMORAI FROM: SUBJECT:	NDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES Dr. Alondra Nelson <i>Judie Villsin</i> Deputy Assistant to the President and Deputy Director for Science and Society Performing the Duties of Director Office of Science and Technology Policy (OSTP) Ensuring Free, Immediate, and Equitable Access to Federally Funded Research
1-11203		5. Example
4. Methods		6. Bonus
		7. Challenges/Limits
		8. "How to" at JPL?
	3. Software	9. Recommendations 10. Reads

This document has been reviewed and determined not to contain export controlled technical data.

2 jpl.nasa.gov

# **Open Paper**

- Publish your paper
  - <u>Good:</u> Drop your submitted manuscript in a NASA repository or an open archive (e.g. Earth and Space Science Open Archive ESSOAr)
  - <u>Better</u>: Choose journals that offer an option for open access fees and pay the fee
  - <u>Best</u>: Choose one of the many journals that only offers open access because NASA/JPL won't pay for optional open access fees

## TL, DR

## Where?

Science Advances Scientific Reports AGU Advances AGU Earth and Space Science EGU journals

## License? Selected by journal

# **Open Data**

- Publish your data
  - <u>Good</u>: pick a license (<u>https://creativecommons.org/choose/</u>) and drag/drop your files to Zenodo (or FigShare)
  - <u>Better</u>: include description of the data
  - <u>Best</u>: adopt popular file formats and metadata conventions in your field (e.g. netCDF Climate and Forecast Conventions)

Free of charge, version controlled, and you get DOIs!

# TL, DR

Where? https://zenodo.org

License? CC-BY

# **Open Software**

- Publish your code
  - <u>Good</u>: pick a license (<u>https://choosealicense.com/</u>) and drag/drop your files to GitHub (or BitBucket)
  - <u>Better</u>: include some description (can be comments within code) and basic installation instructions
  - <u>Best</u>: Describe software dependencies (Linux: aptget, Mac: brew, Windows: Chocolatey; Python: pip), use continuous integration (e.g. Travis), release an image (e.g. Docker)

Free of charge, version controlled, DOIs with Zenodo!

## TL, DR

Where? https://github.com

License? BSD 3-Clause Publish your methods

**Open Methods** 

- <u>Good</u>: short file with copy/paste of how your program was run
- <u>Better</u>: executable scripts
- <u>Best</u>: Jupyter Notebook

It'll help remember how you actually did this!

## TL, DR

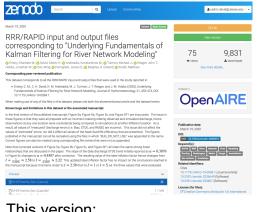
Where? With software

## License? Same as software

# Example

Emery et al. (2020)

### Data



### This version:

### https://doi.org/10.5281/zenodo.3688691 All versions:

https://doi.org/10.5281/zenodo.3688690

Paper

### Software

checked Spiel on Vanis Jpiel John Vanis     Checked Spiel John Vanis     Spiel John Vanis       image     More referentings     Spiel John Vanis       image     Optimise Antoning Spiel     Spiel John Vanis       image     Optise Antoning Spiel     Spiel John Vanis <th>More referred type: a divergence: a divergence: b divergence:</th> <th><ul> <li>* Lode +</li> </ul></th> <th></th>	More referred type: a divergence: a divergence: b divergence:	<ul> <li>* Lode +</li> </ul>	
Interface     Updated authoratio yaar.     Matt month       indigen     Updated authoratio yaar.     Matt mon	trongerse         Uddated authoratio year.         RedAted.End         Removed DISTALL2D year.         RedAted.End         Removed DISTAL2D year.         Removed DISTAL2D year.         Remove	o 🕤 559 commits	
Botompore     Ubdefined authorship year.     But month       Information     Ubdefined authorship year.     But month       Strate year     But month       Dockerline     Ubdefined authorship year.     But month       Instrate genergine Shareship     Ubdefined authorship year.     But month       Inst	Advancement     Advanceme	29 days ago	
gingwere     Opdanted authorskip spart.     But month       a twark poli     Staft om Threinig bies two.     37 dyrs page       CONTINUETOR     Adad datad A. Threinido J.     Harm month       Di Obdantifi     Updated authorskip spart.     But month       BURDET     Buton Marchina J.     Buton Marchina       BURDET     Buton Marchina J.     Buton Marchina       MARMET     Buton Marchina J.     Buton Marchina       MARMET     Buton Marchina J.     Buton Marchina       moid, sparksynamich     Updated authorskip spart.     Buton Marchina       moid, sparksynamich, sparksynamich     Updated authorskip spart.     Buton Marchina       moid, sparksynamich, sparksynamich     Updated authorskip spart.		last month	
Imperformed     Spath own Three (pace hands).     22 days ago       Controller     Added Ammark A Three hands).     10 monthe sign       De Consentie     Updated authorative years:     Bater monthe       DE CDERE     Updated authorative years:     Bater monthe       Gedeentie     Updated authorative years:     Bater monthe       Gedeentie     Updated authorative years:     Bater monthe       Gedeentie     Updated authorative years:     Bater monthe       redd, spathersawerkt     Updated authorative years:     Bater monthe       Replice     Stater monthersawerkt     Imperformed authorative years:     Bater monthersawerkt       Controller     Updated authorative years:     Bater monthersawerkt     Imperfor	a strate synthy a strate synthy b contention b contention b contention b contention b contention b contention c c contention c c contention c c c c c c c c c c c c c c c c c c c	last month	
CONTINUE/TNG     Added Ahmed A. Tavakoja,     10 montina go       Deckente     Updated authorship yara.     Uur montin       MAAME.md     Ramoved MISTALLZ,DR environment sur.     Uur montin       deckartistarpit     Updated authorship yara.     Uur montin       media.prinzega hi     Updated authorship yara.     Uur montin       E     KADME.materine     Jur       C     Line montin     Jur     Jur       C     Line montin     Jur     Jur       D     Line montin     Jur     J	Contribution of the second sec	last month	
Cookersfie     Opdated authorship year.     But moch       UCNE     Opdated authorship year.     But moch       RADAEL-R-     Reveal (RALL) Z/R movement var.     But moch       reveal-stanger     Opdated authorship year.     But moch       reveal-stanger/purgets.hv     Opdated authorship year.     But moch       RADELEN     Stanger     Stanger       Franker     Stanger     Stanger       Franker     Stanger     Stanger       Fraduetstanger     Stanger     Stanger <th>Cockettie Cockettie Cocke</th> <th>27 days ago</th> <th></th>	Cockettie Cocke	27 days ago	
BALANE.med     Updated authorship year.     Uut month       BEADME.md     Removed BETATALIZ DE maivonnent var.     Uut month       okcientestoped     Updated authorship year.     Uut month       reduk_peringvarget.ht     Updated authorship year.     Uut month       reduk_peringvarget.ht     Updated authorship year.     Uut month       reduk_mathurwege.ht     Updated authorship year.     Uut month       reduk_mathurwege.htm     Updated authorship year.     Iut month       reduk_mathurwege.htm     Updated authorship year.     Iut month       reduk_mathurwege.htm     Updated authorship year.     I	Clicity Control Contr	16 months ago	
RAAME.md     Runnood NISTALIZ.DB environment suc:     Luit month       @ dockmitstappin     Updated authorship suc:     Luit month       mold.updatexy.suc:     Luit month       Luit month     Luit	README.md Removed NISTALIZ.DR environment var.     Gebatestarsprif Ustatise authorship yare.     redut spring vary.     RAPME.md  RAPPID  CELESSIONERSTORE  This version:  https://doi.org/10.5281/z	last month	
Ockerzetezymi     Updawie zauthowiło ysaz.     Lut moch.       ordzi, zypierzywarki     Lut moch. <td< th=""><th>b dedeteratory b ded</th><th>last month</th><th></th></td<>	b dedeteratory b ded	last month	
nejidprijota.ywarsh     Opdated authorship yaar.     Batt month       nejidprijota.ywarsh     Opdated authorship yaar.     Batt month       nejida.ywarsh.uh     Opdated authorship yaar.     Batt month       nejida.ywarsh.uh     Opdated authorship yaar.     Batt month       regida.ywarsh.uh     Opdated authorship yaar.     Ital month       regida.ywarsh.uh     Ital month     Ital month <t< th=""><th>relid applinayaeth Uodeled authoritie yeer. relid applinayaeth Uodeled authoritie yeer. relid applinayaeth Uodeled authoritie yeer. relid applinayaeth Uodeled authoritie yeer. RADID RADID RADID This version: https://doi.org/10.5281/z</th><th>last month</th><th></th></t<>	relid applinayaeth Uodeled authoritie yeer. relid applinayaeth Uodeled authoritie yeer. relid applinayaeth Uodeled authoritie yeer. relid applinayaeth Uodeled authoritie yeer. RADID RADID RADID This version: https://doi.org/10.5281/z	last month	
Implif.htmdt_provegach     Opdieled authorskip yaar.     Buit munch.       Implif.htmdt_provegach     Opdieled authorskip yaar.     Buit munch.       Implif.htmdt_provegach     Updieled authorskip yaar.     Buit munch.       Implif.htmdt_provegach     Updieled authorskip yaar.     Buit munch.       Implif.htmdt_provegach     Updieled authorskip yaar.     Buit munch.       Implif.htmdt_provegach     Implif.htmdt     Implif.htmdt       Implif.htmdt_provegach     Implif.htmdt     Implif.htmdt       Implif.htmdt     Implif.htmdt<	medicineal_presetability     Uddited authorable year.       medicineal_presetability <th>last month</th> <th></th>	last month	
Implifying start     Opplified authorship year.     Batt much.       Implifying start     Opplified authorship year.     Batt much.       Implifying start     Implifying start     Implifying start	Implements at     Underland suffering year.       Implements at     Implements at       Implements at     Implements at <th></th> <th></th>		
requirements.et     topatine authority yee:     tet month     RAPIC      RAPIC      Definition      Defin	Consideration of the sufficiency set.  Consideration of		
E RADMLand RAPID Collisisticurus SSSS Kores Exceptions	RAPHEM RAPID CONTRACTOR CONTRACTOR This version: https://doi.org/10.5281/z		
RAPID	RAPID This version: https://doi.org/10.5281/z	last month	
Section States	This version: https://doi.org/10.5281/z	0	
Section States	This version: https://doi.org/10.5281/z		
Send (11) Febru	This version: https://doi.org/10.5281/z		
bulk jaming	This version: https://doi.org/10.5281/z		
	This version: https://doi.org/10.5281/z		
eccurr build persing	This version: https://doi.org/10.5281/z		
	https://doi.org/10.5281/z		
	https://doi.org/10.5281/z		
I his version:			
	All versions:	<u>enodo.376</u>	<u>534</u>
This version: https://doi.org/10.5281/zenodo.376534	All Versions'		
https://doi.org/10.5281/zenodo.376534			

https://doi.org/10.5281/zenodo.593867

This document has been reviewed and determined not to contain export controlled technical data.

jpl.nasa.gov

# Bonus stuff you'll learn to love about open science

### Continuous Integration (Travis CI)

#### docker hub Q Search for great content (e.g., mysq \* Travis Cl 🛞 Dashboard Changelog Documentation Help Explore chdavid/rapid chdavid/rapid 🌣 Manage Repository 🗉 c-h-david / rapid 💭 💷 By chdavid • Updated a month ago Routing Application for Parallel computation of Discharge (RAPID) More options = My Repositories + Current Branches Build History Pull Requests ✓ c-h-david/rrr # 262 20210423 Split one Travis job into two. #186 passed ③ 10 hrs 55 min 24 sec 伦 Cedric H. David -o- dc6e359 12 27 days ago Overview Dockerfile Builds Duration: 27 min 45 sec Tags Finished: 10 days ago () 11 hrs 22 min 50 sec master Split one Travis job into two. ---- #185 passed 🏶 Cedric H. David -o- dc6e359 🖉 27 days ago ✓ c-h-david/rapid # 186 Advanced Image Management ( View preview View all your images and tags in this repository, clean up unused content, recover untagged images. Available for Pro and Team accounts Duration: 10 hrs 55 min 24 sec master Minor reformatting. --- #184 errored ③ 11 hrs 26 min 52 sec Finished: 27 days ago Redric H. David 27 28 days ago -o- 12208c5 12 Sort by Newest ✓ c-h-david/shbaam **# 122** master Updated authorship year. -∽ #183 errored ③ 10 hrs 55 min 5 sec O Duration: 31 min 8 sec 🏶 Cedric H. David 28 days ago -o- fb9a21d 12 Finished: about a year ago 20210423 docker pull chdavid/rapid:20210423 Ē v1.8.0 ③ 10 hrs 7 min 3 sec Added testing for Emery et al. (2020). #182 passed ✓ c-h-david/moshpyt # 8 Last pushed a month ago by chdavid 🏶 Cedric H. David -o- ad2363c 12 i about a year ago Duration: 4 min 24 sec c734c858063h 773.55 MB Finished: 2 years ago linux/amd6/ 25 days ago / 20200424 ③ 9 hrs 56 min 35 sec Added testing for Emery et al. (2020). #181 passed 🏶 Cedric H. David -o- ad2363c 12 about a year ago

Containerization (Docker)

<u>Peace of mind</u>: every update (even minor is fully tested), you can know if you broke anything <u>Less email traffic</u>: "I can't install your code!", "what's wrong with my file?" <u>Faster research</u>: next team member (yourself included) can grab the previous study and build on it

# **Community Challenges and the "Limits of Sharing"**

- Technical training  $\rightarrow$  we need to know/teach how to do this
- Self-perceived inadequacy → "not good enough"
- Documentation  $\rightarrow$  takes time and effort
- Acknowledgment, evaluation, recognition of digital scholarship → "carrot" (citations, metrics, annual reviews, awards)
- Sustainable sharing  $\rightarrow$  open science is <u>not</u> free support
- Keep up  $\rightarrow$  rapidly-evolving tools for open science require time/effort

## Suggested reads

#### commentary

### Of carrots and sticks

to change

#### Jens Kattge, Sandra Díaz and Christian Wirth

Journals and funders increasingly require public archiving of the data that support publications. We argue that this mandate is necessary, but not sufficient: more incentives for data sharing are needed.

collaborations, but stronger incentives

scientific system that it needs to be.

Constraints to data sharing are now more

social than technical. Scientists tend to

in bilateral contexts, but they are often

embrace the opportunity of sharing data

scientific community. The reasons for this

data sets are amalgamated in large collective

making hard-earned data publicly available

opportunities. For example, FLUXNET

workflows allowing globally integrated

analyses of ecosystem-atmosphere

are manifold, but we feel that three are

Three stumbling blocks

particularly prominent.

/ ith the digital revolution, data exchange has become easy in a technical sense. As a result, data that were originally collected for one specific purpose can now be used in different contexts, to answer new scientific questions. Data sharing offers prospects for progress, and not only for data-intensive sciences like remote-sensing. Scientific domains that are typically dominated by numerous small data sets - such as ecology, biodiversity or medicine - stand to benefit, too.

Historically, data sharing was limited by the absence of centralized easily accessible archives for scientific data. Data were usually stored at the research institutions where they were produced, and formats ranged from a centralized institutional digital repository to hand-written field notebooks. Upon publication, the underlying data sets were typically shared via bilateral communication between researchers, and mostly used solely to repeat a given study and verify its results. In such informal structures, metadata are often lacking and data are prone to rapid loss of information content, which makes reuse difficult<sup>1</sup>

As the focus of much research is shifting towards larger-scale questions - such as global biodiversity scenarios, worldwide organismal specialization patterns and

#### Box 1 | Types and examples of data depositories

Generic data depositories, such as PANGAEA (www.pangaea.de) or DRYAD //datadryad.org), compile data sets from a wide range of scientific domains. They guarantee long-term availability of contributed data sets, can ensure the presence of appropriate metadata to some extent and provide an opportunity to make data widely visible and accessible.

Domain-specific data repositories, such as the FLUXNET (http://fluxnet.ornl.gov) database for micro-meteorological eddycovariance measurements or the TRY (www. try-db.org) database for plant traits, cover

continental pandemics - and data collection is understandable — at least as long as the is becoming ever more efficient, the approach measurements have not been sufficiently to storing and disseminating data needs exploited by those who obtained them.

Second, data are context dependent. We argue that data sharing is already Without appropriate contextual information, rewarded with recognition, influence and for example metadata regarding locations, methods and shortcomings, they can easily in terms of citations are overdue. Only if be misinterpreted and misused. Opening the full scientific value of generating and data sets to other researchers means that the disseminating data is acknowledged, will circumstances of collection - known by those data sharing become the integral part of the who performed the measurements - need to be carefully recorded and communicated. The third factor is related to the previous

one: significant effort is often necessary to prepare the data for reuse before they can be made available to the scientific community. Ir addition to contextualising data, formats may have to be adjusted and a point of contact reluctant to release their data to the broader may be necessary in case there are questions In order to overcome these obstacles to voluntary data sharing, public archiving has been made mandatory by many First, high-quality data are hard to obtain. Researchers should expect their fieldwork publishers and funding agencies. However this approach - using a proverbial stick to yield one or several primary publications. to encourage data sharing - has not (yet) led to a broad cultural change in The originality of these publications might e jeopardized if the data are widely available researchers' actions. We therefore advocate before they are published, or if individual complementary incentives - a carrot that

will supplement the stick. synthesis publications. The reluctance towards Big data, many references

There are already incentives for sharing high quality data accompanied by all the relevant contextual information. Benefits to those who readily make their data available includ a narrower range of data, but in turn offer the facilitation of new collaborations and the more intense data curation and networking development of their professional network; researchers can enhance their own original has developed standardized data curation data set by allowing others to access it and contribute; data sharing can also result in joint publications with other groups who use the data, and it may yield data publications and hence citations beyond those of the original papers.

> cultivated as much as they could be, but improvements are under way. Examples include the development of domain-specific data repositories, which explicitly support networking opportunities (see Box 1). But collaborations and networks tend to form

NATURE GEOSCIENCE | VOL 7 | NOVEMBER 2014 | www.nature.com/baburezeoscience

### WORLD VIEW A personal take on events



# Publish your computer code: it is good enough Preely provided working code – whatever its quality – improves programming

benefit in this way

NO EVENSES

NOBODY IS ENTITLED

**TO DEMAND** 

TECHNICAL SUPPORT

FOR FREELY

PROVIDED CODE:

IF THE FEEDBACK

and enables others to engage with your research, says Nick Barnes.

them and now intends to replace its original software with ours.

So, openness improved both the code used by the scientists and the

ability of the public to engage with their work. This is to be expected.

Other scientific methods improve through peer review. The open-

source movement has led to rapid improvements within the software

industry. But science source code, not exposed to scrutiny, cannot

If scientists stand to gain, why do you not publish your code? I have

already discussed misplaced concern about quality. Here are my

It is not common practice. As explained above, this must change in

such as bioinformatics, are already changing.

People will pick holes and demand support and

bug fixes. Publishing code may see you accused of

of fraud. Which is worse? Nobody is entitled to

demand technical support for freely provided code: if the feedback is unhelpful, ignore it.

The code is valuable intellectual property that

belongs to my institution. Really, that little MAT-

LAB routine to calculate a two-part fit is worth

money? Frankly, I doubt it. Some code may have

long-term commercial potential, but almost all

sloppiness. Not publishing can draw allegations

climate science and should do so across all fields. Some disciplines,

responses to some other common excuses.

am a professional software engineer and I want to share a trade secret with scientists: most professional computer software isn't very good. The code inside your laptop, television, phone or car is often badly documented, inconsistent and poorly tested. Why does this matter to science? Because to turn raw data into

published research papers often requires a little programming, which means that most scientists write software. And you scientists generally think the code you write is poor. It doesn't contain good comments, have sensible variable names or proper indentation. It breaks if you introduce badly formatted data, and you need to edit the output by hand to get the columns to line up. It includes a routine written by a graduate student which you never completely understood, and so on. Sound familiar? Well, those things don't matter.

That the code is a little raw is one of the main reasons scientists give for not sharing it with others. Yet, software in all trades is written to be good enough for the job intended. So if your code is good enough to do the job, then it is good enough to release - and releasing it will help your research and your field. At the Climate Code Foundation, we encourage scientists to publish their software. Our experience shows why this is important and how researchers in all fields can benefit. Programs written by scientists may be small

scripts to draw charts and calculate correlations, trends and significance, larger routines to process and filter data in more complex ways. or telemetry software to control or acquire data from lab or field equipment. Often they are an awkward mix of these different parts, glued together with piecemeal scripts. What they have in common is that, after a paper's publica-

tion, they often languish in an obscure folder or are simply deleted. Although the paper may include a brief mathematical description of the processing algorithm, it is rare for science software to be published or even reliably preserved.

from the University of East Anglia (UEA) in Norwich, UK, highlighted the issue, and led the official inquiry to call for scientists to publish code. My efforts pre-date the UEA incident and grew from work in 2008 based on software used by NASA to report global temperatures. Released on its website in 2007, the NASA code was messy and proved difficult for critics to run on their own computers. Most did not seem to try very hard, and nonsense was written about fraud and conspiracy With other volunteers. I reserve

understand and run. All software has bugs, and Discuss this article we found a number of minor problems, which online at. had no bearing on the results. NASA fixed go.sature.com/ef3hsl

lishing; those who refuse are blocking progress. It is too much work to polish the code. For scientists, the word publication is totemic, and signifies perfectionism. But your papers need not include meticulous pages of Fortran; the original code can be published as supplementary have all called for transparency. To make it happen, they have to be

> you should fix that anyway. # SEENEWS FEATURE P. 775 Sheffield S17 4DL, UK.

#### **@AGU** PUBLICATIONS

#### Earth and Space Science

#### RESEARCH ARTICLE 10.1002/2015FA000142

#### A decade of RAPID—Reflections on the development of an open source geoscience code

#### Cédric H. David<sup>1,2</sup>, James S. Famiglietti<sup>1,2,3</sup>, Zong-Liang Yang<sup>4</sup>, Florence Habets<sup>5</sup>, and Special Section: Geoscience Papers of the David P. Maidmont<sup>6</sup>

University of California, Irvine, California, USA, <sup>3</sup>Department of Earth System Science, University of California, Irvine, Key Points: · A reflection on the open source California, USA. <sup>4</sup>Department of Geological Sciences, Jackson School of Geosciences, University of Texas at Austin, Austin development of geoscience codes is Texas, USA, <sup>5</sup>UMR 7619 METIS, CNRS, UPMC, Paris, France, <sup>6</sup>Center for Research in Water Resources, University of Texas at presented Austin, Austin, Texas, USA Sharing can be broken down into three phases opening, exposing, consolidating • Free online services facilitate sharing

and allow for further academic credit

Correspondence to: C. H. David, cedric david@inl.nasa.oc/

David, C. H., J. S. Famiglietti, Z.-L. Yang, F. Habets, and D. R. Maidment (2016), A

©2016 The Authors

This is an open access article under the

Attribution-NonCommercial-NoDerive

License, which permits use and distr

bution in any medium, provided the

non-commercial and no modification

or adaptations are made

original work is properly cited, the use is

terms of the Creative Commons

decade of RAPID-Reflections on the development of an open source geoscience code, Earth and Space Science, 3, 226-244, doi:10.1002/

Received 13 OCT 2015 Accepted 22 MAR 2016 Accepted article online 7 APR 2016 Published online 19 MAY 2016

Abstract Earth science increasingly relies on computer-based methods and many government agencies now require further sharing of the digital products they helped fund. Earth scientists, while often supportive of more transparency in the methods they develop, are concerned by this recent requirement and puzzled by its multiple implications. This paper therefore presents a reflection on the numerous aspects of sharing code and data in the general field of computer modeling of dynamic Earth processes. Our reflection is based on 10 years of development of an open source model called the Routing Application for Parallel Computation of Discharge (RAPID) that simulates the propagation of water flow waves in river networks. Three consecutive but distinct phases of the sharing process are highlighted here: opening, exposing, and consolidating. Each

<sup>1</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA, <sup>2</sup>Center for Hydrologic Modeling

one of these phases is presented as an independent and tractable increment aligned with the various stages of code development and justified based on the size of the users community. Several aspects of digital scholarship are presented here including licenses, documentation, websites, citable code and data repositories, and testing. While the many existing services facilitate the sharing of digital research products,

digital scholarship also raises community challenges related to technical training, self-perceived inadequacy, community contribution, acknowledgment and performance assessment, and sustainable sharing

#### 1 Introduction

Driven by the need to understand Earth's dynamic climate, geoscientists have dedicated much effort to creating numerical models of the major components of the climate system and to analyzing their outputs. Early modeling studies date back to the 1950s and include simulations of the Earth's atmosphere [Phillips, 1956], oceans [Bryan and Cox, 1967], land [Manabe, 1969], and rivers [Miller et al., 1994]. Decades later, computer modeling and data-intensive analysis have become key elements upon which modern climate science has been built [e.g., Intergovernmental Panel on Climate Change, 2013], and numerous geoscientists therefore dedicate considerable research energy to such endeavors. Computer-assisted research is equally ubiquitous in the broad scientific community, such that some have argued that computer modeling and data-intensive science be considered legitimate pillars of science, hence joining experimental science and theoretical science [Bell, 1987; Bell et al., 2009; Hey et al., 2009; Hey, 2010; Hey and Payne, 2015], although such a view is not without its critics [Vardi, 2010a, 2010b]. Nevertheless, computer modeling and analysis are now integral parts of many geoscience investigations

The recent mandate [Holdren, 2013] requesting that the direct results of federally funded scientific research in the U.S. be made further accessible-including availability of digital data-has spurred much discussion in the scientific community. Kattge et al. [2014] argued that while data sharing is necessary, associated hurdles subsist, and proper means of acknowledgment (i.e., citations) are needed so that scientists can benefit from the added burden. This argument was further supported by the survey of Kratz and Strasser [2015]. Others have also suggested that the computer codes used to generate or to analyze data are equally important and should hence be made similarly accessible [Nature, 2014; Nature Geoscience, 2014]. Prior to the recent mandate, Barnes [2010] had already advocated for sharing computer code so that-like any other scientific method-code development could benefit from the peer review process. Additionally, the description of computations using only natural language or equations has inherent ambiguities that have unpredictable effects on results: hence, access to the source code is essential to reproducing the central findings of studies

A DECADE OF RARID

226

https://doi.org/10.1002/2015EA000142

### https://doi.org/10.1038/467753a

11/4/22

This document has been reviewed and determined not to contain export controlled technical data.

ipl.nasa.gov

the value lies in your expertise. My industry has a name for code not backed by skilled experts: IS UNHELPFUL, **IGNORE IT** abandonware. Institutions should support pub-

> rmation, available from an institutional or journal website. I accept that the necessary and inevitable change I call for cannot be made by scientists alone. Governments, agencies and funding bodies

prepared to make the necessary policy changes, and to pay for training, workshops and initiatives. But the most important change must come in the attitude of scientists. If you are still hesitant about releasing your code, then ask yourself this question: does it perform the algorithm you describe in your paper? If it does, your audience will accept it, and maybe feel happier with its own efforts to write programs. If not, well,

Nick Barnes is director of the Climate Code Foundation, e-mail: nb@climatecode.org

14 OCTOBER 2010 | VOL 467 | NATURE | 753

These benefits have perhaps not been

© 2014 Macmillan Publishers Limited. All rights reserved

https://doi.org/10.1038/ngeo2280

#### exchange across all measurement sites: the TRY database facilitates outlier detection. duplicate identification and gap-filling of missing data: GBIF the Global Biodiversity Facility (www.gbif.org) compiles occurrence data for all kinds of species and has developed highly efficient algorithms to identify and potentially correct mistakes in geo-locations



Last year's global fuss over the release of climate-science e-mails

the software to make it easier for non-experts to ONATURE.COM

© 2010 Macmillan Publishers Limited. All rights reserved



jpl.nasa.gov