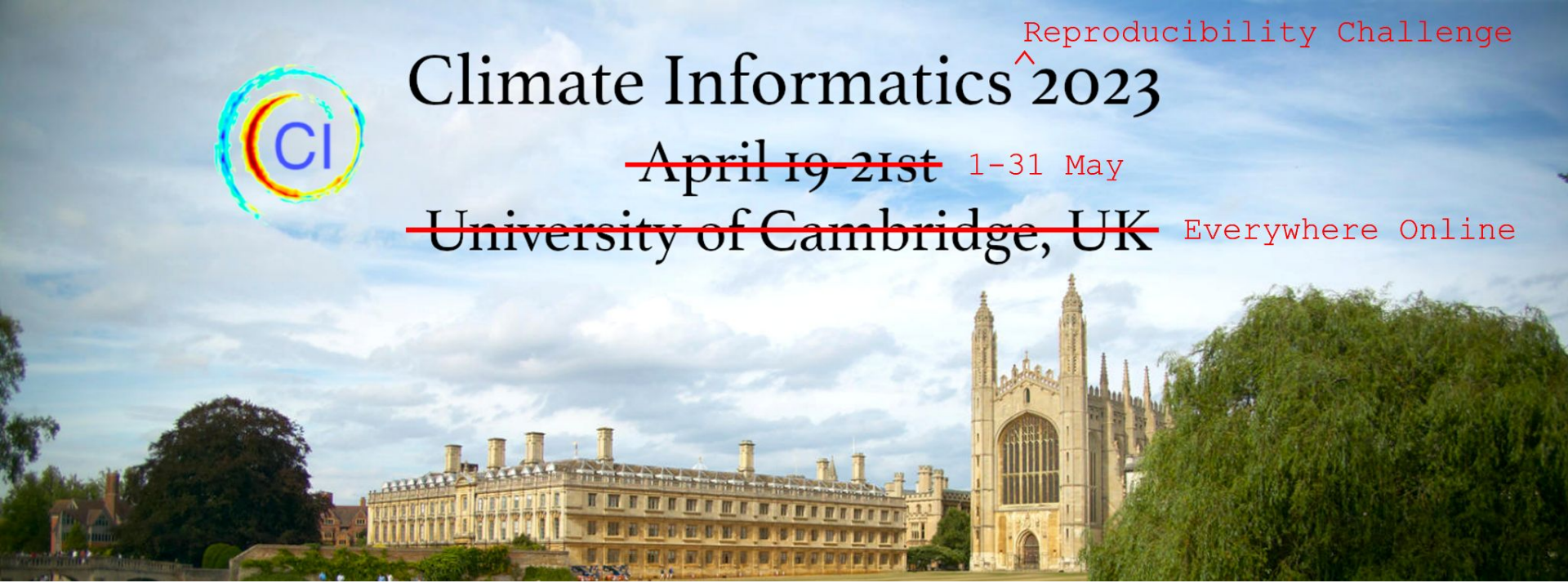


Kickoff
2 May 2023

CI2023 Reproducibility Challenge Planning Committee
Alejandro Coca-Castro, Andrew Hyde, Andrew McDonald, Anne Fouilloux

bit.ly/ci-2023-rc-eds

Agenda

- 14:00-14:50 Welcome
  - Climate Informatics (Andrew McDonald)
  - Environmental Data Science Journal (Andrew McDonald)
  - Environmental Data Science Book (Alejandro Coca-Castro)
  - Logistics & Communications (Alejandro Coca-Castro)
  - Timeline & Schedule (Anne Fouilloux)
  - Q&A
- 14:50-15:00 Break
- 15:00-15:50 Technology Overview
  - C-SCALE (Anne Fouilloux)
  - JupyterHub (Anne Fouilloux)
  - EDS book submission demo (Alejandro Coca-Castro)
  - Q&A
- 15:50-16:00 Break
- 16:00-16:30 Speed Networking (optional)

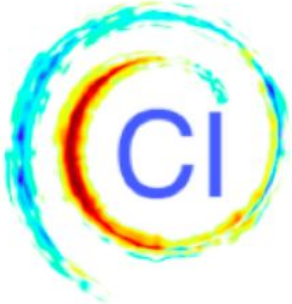
Welcome

Climate Informatics

An open community at the intersection of climate science and data science.

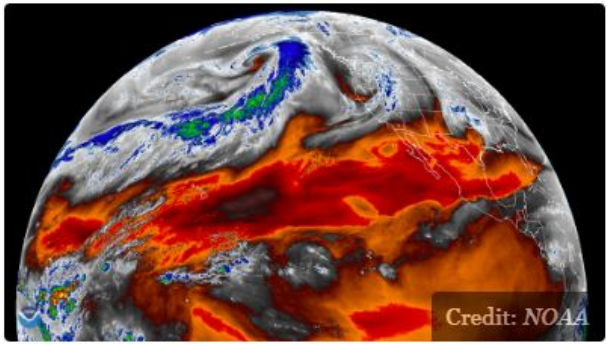
Join the Community

Climate Informatics is an open community interested in research combining climate science with approaches from statistics, machine learning and data mining. Through the annual conference series and through this community website we hope to stimulate discussion of new ideas, foster new collaborations, grow the community, and thus accelerate discovery across disciplinary boundaries.



Climate Informatics

An open community at the intersection of climate science and data science.



Climate Centered

Climate Informatics is centered around the discovery of climate sciences with state-of-art methods in statistics, machine learning, and data mining.

Community Driven

The community is driven by active members organizing annual conference series, hackathons, and other events to promote advancement in the field.

Collaborative Innovation

We promote collaborations and support early-career scientists by bringing experts across disciplines together through various events and online forums.

Claire Monteleoni
University of Colorado Boulder

Gavin Schmidt
NASA

Claire Monteleoni
University of Colorado Boulder

Jakob Runge
German Aerospace Center (DLR)

Eniko Szekely
Swiss Data Science Center

Anastase Alexandre Charantonis
ENSIIE

2011
New York

2012-2018
Boulder

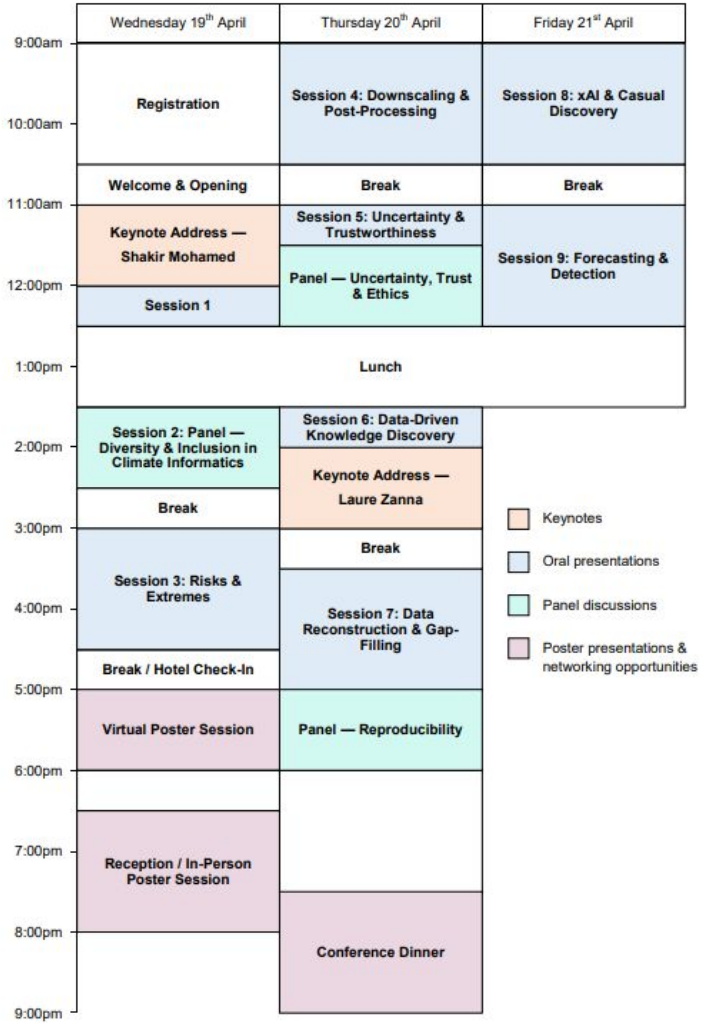
2019
Paris

2020
Oxford

2022
Asheville

2023
Cambridge

Programme at a Glance



University of Cambridge
Institute of Computing for Climate Science

Schmidt Futures

British Antarctic Survey
Natural Environment Research Council

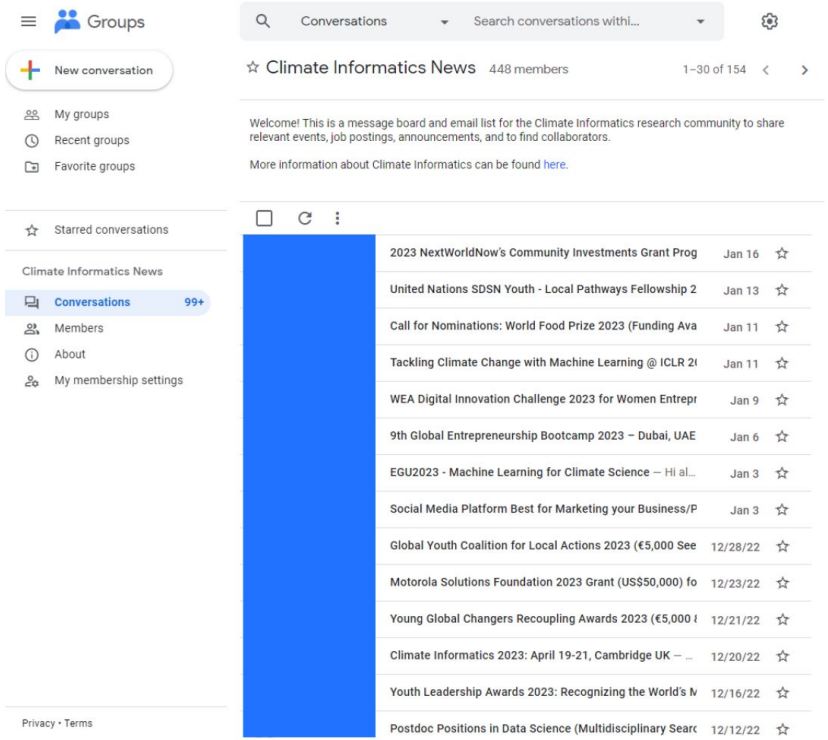
aws

NCICS
North Carolina Institute for Climate Studies

The Alan Turing Institute

Cambridge University Press & Assessment

@Climformatics
https://twitter.com/Climformatics



https://groups.google.com/g/climate-informatics-news

Environmental Data Science

Access: Subscribed, Past subscription | Open access

ISSN: 2634-4602 (Online)

Editor: Claire Monteleoni *University of Colorado Boulder, USA*

Editorial board

Environmental Data Science is an open access journal dedicated to the use of data-driven approaches to understand environmental processes - including climate change - and aid sustainable decision-making. The data and methodological scope is defined broadly to encompass artificial intelligence, machine learning, data mining, computer vision, econometrics and other statistical techniques.

EDS is a venue for application and methods papers, whether they relate to the geosphere (the solid earth and its processes), cryosphere (e.g. ice, snow, permafrost and tundra), biosphere (ecology), hydrosphere (oceans and fresh water, including the water cycle) or atmosphere (e.g. meteorology, climatology). It also welcomes work that shows how data science can inform societal responses to environmental problems (such as climate change, air quality, energy, natural resources and land use).

EDS promotes open data and data re-use - through data papers that describe valuable environmental data sets - and publishes shorter position papers relevant to the journal's scope.

Open access

View the editorial board

Environmental Data Science Journal Editorial Board (May 2023)

Editor-in-Chief
Claire Monteleoni, *University of Colorado Boulder, USA*

Editors
Arindam Banerjee, *University of Illinois Urbana-Champaign, USA*
Julien Brajard, *Nansen Center (NERSC), Norway / Sorbonne University, France*
Anastase Alexandre Charantonis, *ENSIIE, Évry, France*
Chen Chen, *Center for Climate Research, Singapore*
Felix Creutzig , *Technical University of Berlin, Germany*
Bistra Dilkina, *University of Southern California Viterbi School, USA*
Auroop Ganguly, *Northeastern University, USA*
Sophie Giffard-Roisin, *Grenoble Alpes University, France*
Dorit Hammerling, *Colorado School of Mines, USA*
Lester Mackey, *Stanford University & Microsoft Research, USA*
Miguel Mahecha, *Leipzig University, Germany*
Sarvapali Ramchurn, *University of Southampton, UK*
Jakob Runge, *German Aerospace Center, Germany*
Eniko Székely, *Swiss Federal Institute of Technology Lausanne, Switzerland*
Duncan Watson-Parris, *University of California San Diego, USA*

Communications Editor
Ricardo Barros Lourenço, *McMaster University, Canada*

Advisory Board
Gordon Blair, *Lancaster University, UK*
Imme Ebert-Uphoff, *Colorado State University, USA*
Veronika Eyring, *German Aerospace Centre, Germany*
Vipin Kumar, *University of Minnesota, USA*
Doug Nychka, *Colorado School of Mines, USA*
Emily Shuckburgh, *University of Cambridge, UK*
Bhekisipho Twala, *Tshwane University of Technology, Pretoria, South Africa*

ENVIRONMENTAL
DATA SCIENCE

Why choose this journal?

Five reasons to contribute to Environmental Data Science

ENVIRONMENTAL DATA SCIENCE

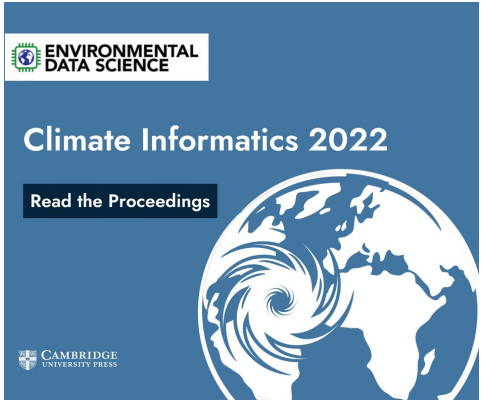
1. Submit your work to a dedicated venue for the use of data-driven methods for the environment

2. EDS gives you quality feedback from methods and domain experts

3. Make your work Open Access (OA) to enhance its dissemination and impact

4. Enrich your article with Open Data and Open Materials

5. Help us build a growing research community using data-driven and open science for the environment



Climate Informatics 2022

Climate Informatics 2022

EDS is pleased to partner with Climate Informatics (CI): an open community at the intersection of climate science and data science. This partnership began in 2022, the 11th International Conference on Climate Informatics, which was held in Asheville, North Carolina, USA on May 19-23. Authors of full papers accepted into CI 2022 had the option of publishing in EDS, the articles being accepted in EDS on the strength of the CI review process. Authors of abstracts presented at the conference were also encouraged to submit to EDS. CI 2022 papers published to date can be found below.

EDS would like to thank Yuhan (Douglas) Rao (North Carolina Institute for Climate Studies) and Chen Chen (Centre for Climate Research Singapore) of the Organizing Committee and Claire Monteleoni (University of Colorado Boulder, Climate Informatics co-Founder and EDS Editor-in-Chief) for facilitating this partnership and the reviewers for committing their time to assessing the papers.

For more information about the community, follow @Climformatics on Twitter and visit the Climate Informatics website.

Articles

Article
A climatic random forest model of agricultural insurance loss for the Northwest United States
Erich Seamon, Paul E. Gessler, John T. Abatzoglou, Philip W. Mote, Stephen S. Lee
Environmental Data Science, Volume 1

Article
Week-ahead solar irradiance forecasting with deep sequence learning
Saumya Sinha, Bri-Mathias Hodge, Claire Monteleoni
Environmental Data Science, Volume 1

Article
Sea surface height super-resolution using high-resolution sea surface temperature with a subpixel convolutional residual network
Théo Archambault, Anastase Charantonis, Dominique Béréziat, Carlos Mejia, Sylvie Thiria
Environmental Data Science, Volume 1

Article
A Gaussian process state-space model for sea surface temperature reconstruction from the alkenone paleotemperature proxy
Taehee Lee, Jun S. Liu, Charles E. Lawrence
Environmental Data Science, Volume 1

Article
Deep prior in variational assimilation to estimate an ocean circulation without explicit regularization
Arthur Filoche, Dominique Béréziat, Anastase Charantonis
Environmental Data Science, Volume 2

Article
Long-term stability and generalization of observationally-constrained stochastic data-driven models for geophysical turbulence
Ashesh Chattopadhyay, Jaideep Pathak, Ebrahim Nabizadeh, Wahid Bhimji, Pedram Hassanzadeh
Environmental Data Science, Volume 2

@EnvDataScience
https://twitter.com/EnvDataScience

Environmental Data Science book

A **living**, **free** and **open** online resource to **showcase** and **support** the **publication** of data, research and open-source tools for **collaborative**, **reproducible** and **transparent** **Environmental Data Science**

CC-BY image by Turing Way and Scriberia, CC-BY 4.0, EDS book community, @eds_book

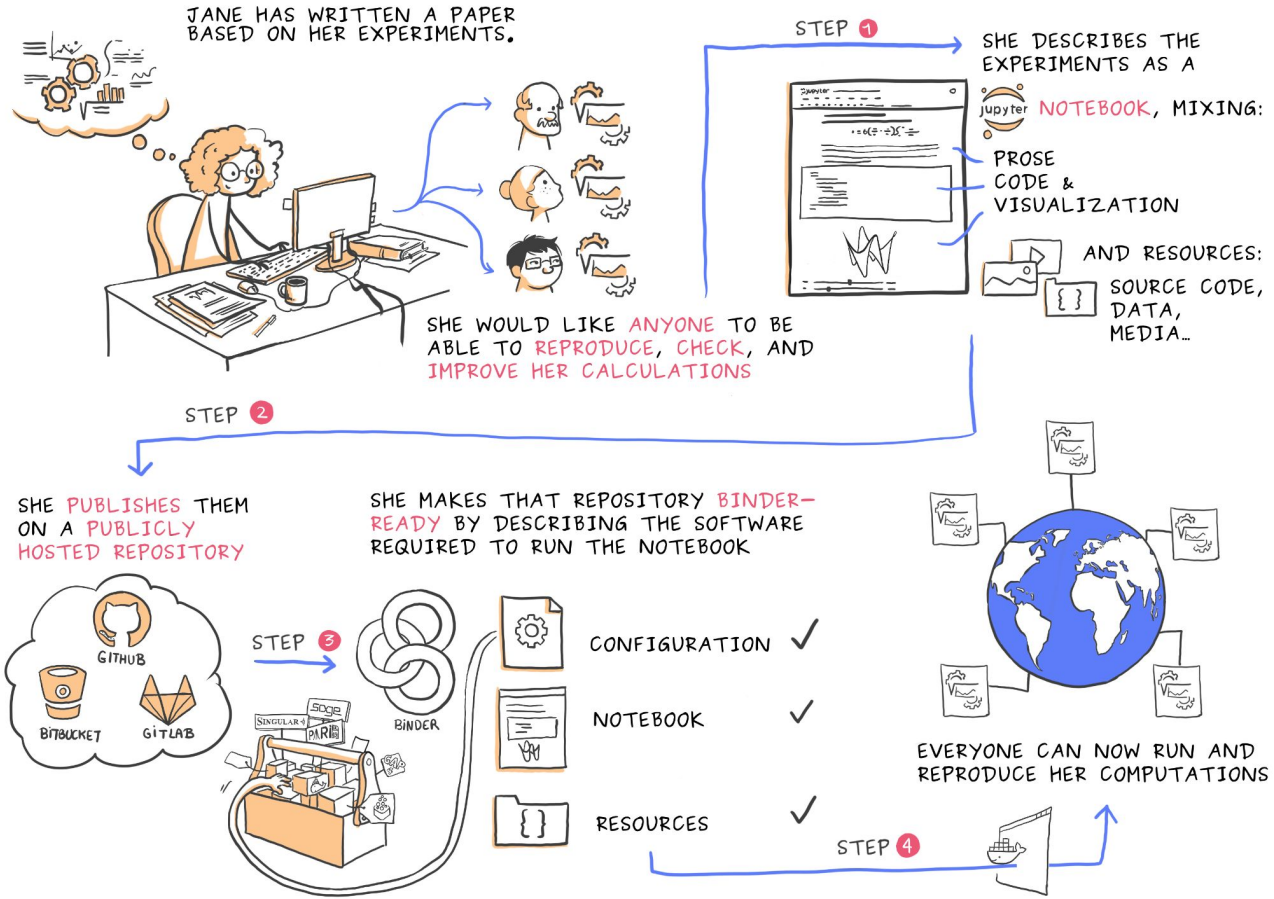
Mission

Educate and leverage good scientific software and data management practices among environmental scientists through peer-reviewed findable, accessible, interoperable and reusable (FAIR) executable notebooks

Vision

Environmental scientists work collaboratively to demonstrate and communicate their science through FAIR executable notebooks and have gained significant skills to publish in notebook-based scholarly publication systems

CC-BY 4.0, EDS book community, @eds_book



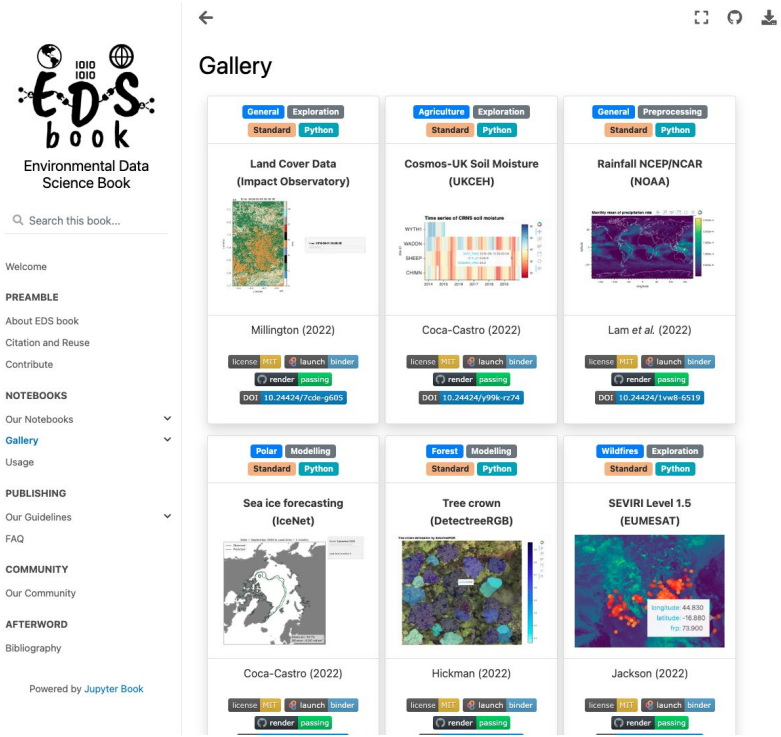
Juliette Taka, & Nicolas M. Thiéry. (2018). Publishing reproducible logbooks explainer comic strip. Zenodo. https://doi.org/10.5281/zenodo.4421040

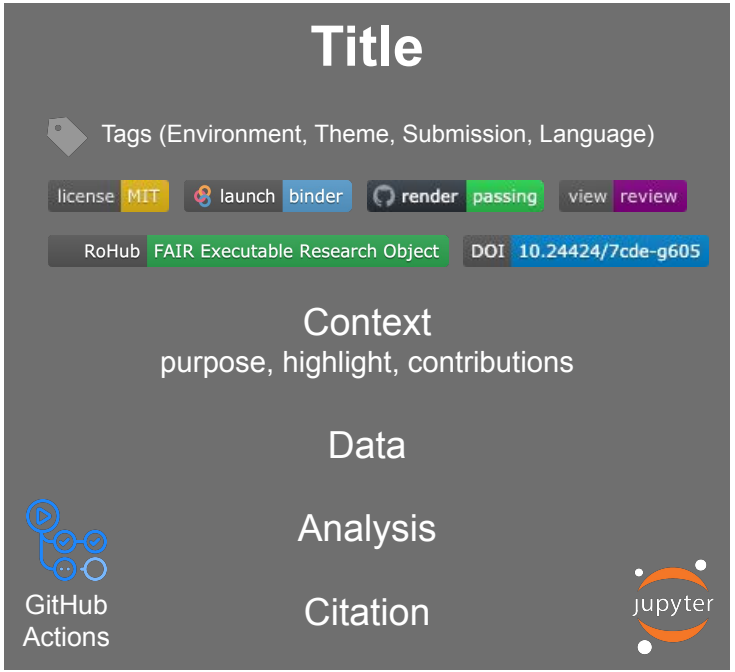
Publishing process: Stages and Roles

Stage	Where in GitHub	Authors	Reviewers	Editors-in-Chief	Editors	Community
Notebook idea	EDS book repo	✅		✅		⭕
Preparation	Notebook repo	✅		✅		
Prereview and Review* (similar to JOSS/JOSE)	EDS book & notebook repos	✅	✅	✅	✅	
Post-print	EDS book & notebook repos	✅	⭕	✅		
Publication	EDS book repo	⭕		✅		
Post-publication	Notebook repo	⭕		✅		✅

CC-BY 4.0, EDS book community, @eds_book

Gallery with FAIR notebooks





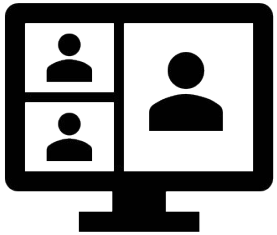
netlify

CC-BY 4.0, EDS book community, @eds_book

Activities in next months

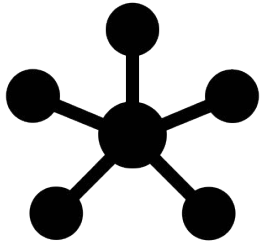
Inform and monitor software stacks health across languages & open-source communities *(Python, Julia, R)*

Community Chapter > Scientific Ecosystems (Issue #154)
e.g. Holoviz Project Dashboard



Joint community calls with The Turing Way
`#environmental-sustainability`
and more co-working activities (maintainers)

Community, 1st/3rd Wednesdays
Co-working, 2nd Friday



Partnerships with environmental research networks, open education, OSS developers and publishing ecosystem

NASA TOPS / Notebooks Now!
Climate Informatics
JupyterCon23

Play a part in EDS book version 0.1.0

CC-BY 4.0, EDS book community, @eds_book

Big thanks!

Open Research Communities

PANGEO

PROJECT PYTHIA

HoloViz

jupyter

The Turing Way

Open Life Science

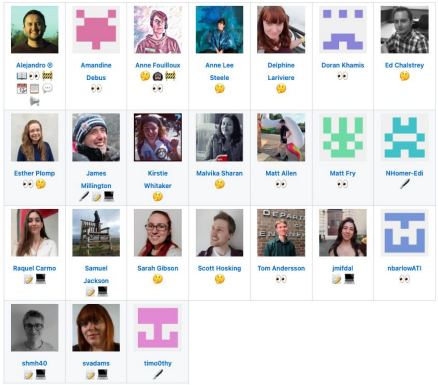
EDS book community

8 Authors

8 Reviewers

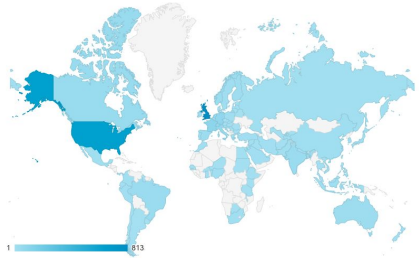
+Advisors/Contributors

10+ participants in community/co-working sessions



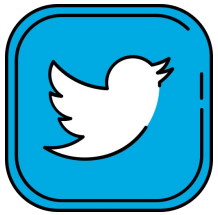
~2300 users

(visits since Dec'21)



Contribution and suggestions are welcome

Follow us!



@eds_book

@EDSbook@fosstodon.org

Visit us!

https://edsbook.org

zenodo.org/communities/
the-environmental-ds-community

Star us!

alan-turing-institute/
environmental-ds-book

'This work was supported by Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1 and EPSRC Grant EP/W006022/1, particularly the "AI for Science" theme within those grants & The Alan Turing Institute'.

Logistics and Communications

Code of Conduct (EDS Book CoC) & GDPR Statement

Communication channels

- Zoom (virtual meetings)
- Slack
  - General channel, **#6-reproducibility-challenge**
  - Teams private channel, **#repro-challenge-team-x**
  - Help channel, **#repro-challenge-help**
- Twitter, @Climformatics

Notes (HackMD)

- Public boards > General Channel
- Teams Notes > Team Private Channel

Timeline

Challenge registration
Registration closes on Friday 28 April @23:59 AOE
Teams & Project assignments
Teams announcement 30 April
Reproducibility challenge Teams' work
Challenge begins on 1 May @ 00:00 AOE and ends on 31 May @ 23:59 AOE
Peer review begins 15 May @ 00:00 AOE
Results
Results announced 15 June
Publication to EDS Book
Submissions are integrated to the EDS Book throughout the summer.

Team and Paper Assignments

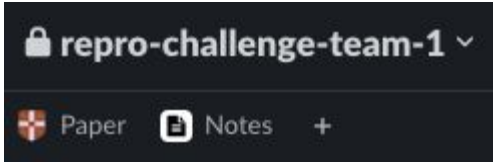
Teams announcement, 30 April noon UTC+1

7 teams, 2-4 members per team

Outcome: A peer-reviewed Jupyter notebook published in EDS book

Private slack channel to internal team communications

+ bookmarks to the paper and notes



		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

CC-BY image by the Turing Way

Teams' work

Notebook first working version, 11 May 23:59 AOE

Peer-review, 15-19 May

Catch Up Presentations, 26 May

Team Presentations, 29 May

Group Submission, 31 May 23:59 AOE

Talk with the Expert

Week 1

Open Infrastructure x1

Sebastian Luna-Valero - EGI Foundation

Week 2

Programming Languages x3

R, Cesar Luis Aybar - rOpenSci
Python, Brian Rose - Pythia Project
Julia, Rafael Schouten - JuliaGeo

Week 3

Reproducible and Collaborative Research x2

Reproducible Publications (TBD),
Collaborative Data Science (The Turing Way)

Week 4

Open Access Journals x1

Cambridge University Press and Assessment

40 - 50 min OPTIONAL:

10 min slides

20 min demo

10-20 min Q&A

Socials

Speed Networking

Today 16:00 UTC+1

Clinics & Coffee Drop-in

Every Friday 15:00-16:00 UTC+1

Final Goodbye and Closure

29 May 16:00-16:45 UTC+1

Schedule

Climate Informatics 2023 Reproducibility Challenge Schedule: 1st - 31st May 2023

EVENT WEEK 1 | Welcome, Teams work & Talk with the Expert on Open Infrastructure

Monday 1			Tuesday 2			Wednesday 3			Thrusday 4			Friday 5		
Time	Activity	Who	Time	Activity	Who	Time	Activity	Who	Time	Activity	Who	Time	Activity	Who
9:00 - 17:00	Introductions at CI2023 slack channel	Participants	14:00 - 15:00	Welcome	Alejandro Coca-Castro, Andrew Hyde, Andrew McDonald, Anne Fouilloux	Anytime	Group work	Teams	Anytime	Group work	Teams	15:00 - 16:00	Clinics & Coffee Drop-in	Alejandro Coca-Castro
			15:00 - 16:00	Training: Technologies Overview	Alejandro Coca-Castro, Anne Fouilloux							16:00 - 17:00	Talk with the Expert - Open Infrastructure	Sebastian Luna-Valero - EGI Foundation
			16:00 - 16:30	Speed Networking	Optional Social							Anytime	Group work	Teams

EVENT WEEK 2 | Teams work & Talk with Experts on Programming Languages

Monday 8			Tuesday 9			Wednesday 10			Thrusday 11			Friday 12		
Time	Activity	Who	Time	Activity	Who	Time	Activity	Who	Time	Activity	Who	Time	Activity	Who
09:00 - 10:00	Talk with the Expert - R programming	Cesar Luis Aybar - rOpenSci	Anytime	Group work	Teams	Anytime	Group work	Teams / Reviewers	Anytime	Group work	Teams	09:00 - 10:00	Talk with the Expert - Julia programming	Rafael Schouten - JuliaGeo
16:00 - 17:00	Talk with the Expert - Python programming	Brian Rose - Pythia Project										15:00 - 16:00	Clinics & Coffee Drop-in	Andrew McDonald
Anytime	Group work	Teams										Anytime	Group work	Teams

EVENT WEEK 3 | Peer-review & Talk with Experts on Reproducible and Collaborative Research

Monday 15			Tuesday 16			Wednesday 17			Thrusday 18			Friday 19		
Time	Activity	Who	Time	Activity	Who	Time	Activity	Who	Time	Activity	Who	Time	Activity	Who
16:00 - 17:00	Talk with the Expert - Reproducible Publications	Speaker to be defined	Anytime	Group work	Teams / Reviewers	Anytime	Group work	Teams / Reviewers	Anytime	Group work	Teams / Reviewers	15:00 - 16:00	Clinics & Coffee Drop-in	Anne Fouilloux
Anytime	Group work	Teams / Reviewers										16:00 - 17:00	Talk with the Expert - Collaborative Data Science	Speaker to be defined - The Turing Way
												Anytime	Group work	Teams / Reviewers

EVENT WEEK 4 | Teams work, Talk with the Expert on Open Access Journal Publishing & Presentations Mock-up

Monday 22			Tuesday 23			Wednesday 24			Thrusday 25			Friday 26		
Time	Activity	Who	Time	Activity	Who	Time	Activity	Who	Time	Activity	Who	Time	Activity	Who
16:00 - 17:00	Talk with the Expert - Open Access Journals	Speaker to be defined - Cambridge University Press and Assessment	Anytime	Group work	Teams	Anytime	Group work	Teams / Reviewers	Anytime	Group work	Teams	15:00 - 16:00	Clinics & Coffee Drop-in	Alejandro Coca-Castro
Anytime	Group work	Teams										16:00 - 17:00	Catch Up Presentations	Teams

EVENT WEEK 5 | Share-out and submissions

Monday 29			Tuesday 30			Wednesday 31		
Time	Activity	Who	Time	Activity	Who	Time	Activity	Who
15:00 - 16:00	Team Presentations	Teams	Anytime	Group work	Teams	23:59 AOE	Group Submission	Teams
16:00 - 16:45	Final Goodbye and Closure	Teams and Organisers						
16:45- 17:00	Tie up loose ends	Teams						

All times listed below are UTC +1 (British Summer Time).
You might want to consult this Time Zone Map to figure out times in your location.

We suggest teams allocate between 2 to 3 hours per day for preparing their submissions.