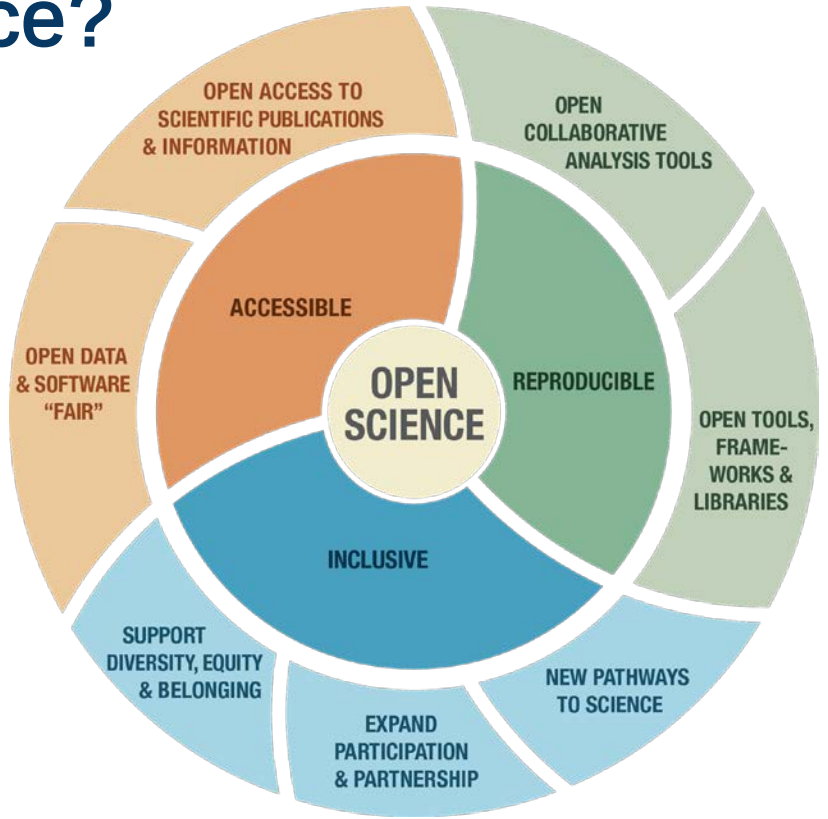
Open-Source Science Initiative

Kevin Murphy, Chief Science Data Officer SMD
Katie Baynes, Deputy Chief Science Data Officer SMD
Dr. Steve Crawford, Science Data Officer SMD
Dr. Chelle Gentemann, TOPS Science Lead
Amy (Uyen) Truong, Chief Science Data Office Coordinator
Christian Reyes, OSSI Coordinator
Yvonne Ivey, TOPS Project Manager
Cyndi Hall, TOPS Community Coordinator
Isabella Martinez, TOPS Curriculum Coordinator
Karla Mastracchio, TOPS Communication Strategy
Yaitza Luna-Cruz, OSSI/TOPS Science Coordinator
Elena Steponaitis, OSSI/TOPS Science Advisor

What is Open Science?

A **collaborative culture enabled by technology** that empowers the **open sharing of data, information, and knowledge** within the **scientific community and the wider public** to accelerate scientific research and understanding.



Science should be…

Transparent
scientific process and results should be visible, accessible, and understandable

Accessible
data, tools, software, documentation, and publications should be accessible to all (FAIR)

Inclusive
process and participants should welcome participation by and collaboration with diverse people and organizations

Reproducible
reproducible by members of the community

Open-Source Science is NASA's method to put Open Science into practice.

- **Open** the entirety of the scientific process, ***from start to finish***
- **Broaden** community involvement in the scientific process
- **Increase** accessibility of data, software, & publications
- **Facilitate** inclusion, transparency, and reproducibility of science

Why Now?

We **now** have the tools to make open science a reality. Advances in technology have created accessible, reproducible, inclusive science at a scale not possible a few years ago.

There is national and global momentum for the move to open science.

Equal and open access benefits the public

Open-Source Science Initiative

Unlocking the full potential of a more equitable, impactful, efficient, scientific future

Policy development, education, compliance tools
Updating NASA policies on scientific information to better enable the activation of open science

Core Services for Science Discovery
Developing core data and computing services to enable open science

ROSES Elements
Supporting open-source software, tools, frameworks, libraries, platforms, and training with over $5 million dollars in grants

Community Building & Partnerships - Transform to Open Science (TOPS)
Accelerating adoption of open science

Advancing Science Requires the *Sharing* of Information

SPD-41 is the NASA SMD Information Policy.

SPD-41 brings together existing NASA and Federal guidance.

It applies to all SMD-funded activities related to producing scientific information.

- SPD-41: The Science Information Policy - https://go.usa.gov/xtNTJ
- Science Information Policy Website - https://go.usa.gov/xtNTt

Feedback on proposed additions to SPD-41 were due by **March 4, 2022**

- An update to SPD-41 will be released no earlier than June 2022

How we share information matters - it affects the impact, the transparency, the reproducibility, and the accessibility of research.

What is the current policy?

Data

Scientific data shall be made publicly available with a clear, open, and accessible data license no later than the publication of the research.

Mission data shall be openly available with no period of exclusive access.

Software

Research software should be publicly available no later than the publication of the research and assigned a permissive software license.

Publications

Manuscripts versions of as-accepted manuscripts shall be deposited in a NASA repository and made publicly available within 12-months.

Mission publications shall additionally be made publicly available at the time of their publication.

What are the new proposed changes?

Data

Scientific data should be FAIR and shall be made publicly available with a clear, open, and accessible data license no later than the publication of the research, and be citable.

Mission data shall be openly available with no period of exclusive access.

Software

Research software shall be publicly available no later than the publication of the research, assigned a permissive software license, and be citable.

Mission software shall additionally be developed openly in a publicly accessible, version-controlled platform that allows for contributions and engagement from the community.

Publications

Manuscripts versions of as-accepted manuscripts shall be deposited in a NASA repository and made publicly available within 12-months. Publishing as open access is supported and posting preprints is encouraged.

Mission publications shall additionally be made publicly available at the time of their publication.

Science workshops and meetings shall be open to broad participation and documented in public repositories.

Open science activities will be considered in reviews of proposals.

Core Services for Science Discovery

Proposal

SMD provides **common directorate capabilities**. Develops core data and computing services to be used as building blocks by divisions and the open science community (to be complete within **three years**).

Objectives

Divisions develop and operate division-specific requirements (**missions and science capabilities**) within SMD core systems.

Meet open source science goals in the Data and Computing Strategy for SMD and requirements in SPD-41.

Reduce cloud environment **development duplication and barriers** to speed of adoption.

Improve computing infrastructure to seamlessly provide access to high-performance computing and cloud resources while reducing cybersecurity risk.

ROSES22 Updates

	Title	Description	Details
F.7	Support for Open Source Tools, Frameworks, and Libraries	Support and maintain open sources tools, frameworks, and libraries that are significantly used by the SMD community	• $2M awarded in ROSES20 to 8 programs • Selection rate of 13% • Once every 3 years
F.8	Supplemental Open Source Software Awards	Supplemental award to encourage the conversion of legacy software to open source	• $200K awarded in ROSES20 to 6 awards • Selection rate of 100% • Yearly, $250K available, rolling deadline
F.14	Transform to Open Science Training	**TOPS** training element primarily solicits proposals for the development training material and the execution of one day meetings, workshops, and summer schools to advance open science literacy	• Budget of $4.5M per year with awards for 3 years • To be released no earlier than April • Once every three years
F.15	High Priority Open-Source Science	SMD seeks proposals to support OSSI and that will advance the goals of TOPS. This includes supporting innovative open source tools, software, frameworks, data formats, and libraries that will have a significant impact to the SMD science community	• Budget ~$1M • Yearly, rolling deadline • Recommended size will be 1 year, $50-100K • Priorities will be related to different Open Source Science objectives • To be released no early than April
F.16	Supplement for Software Platforms	Supplemental support to existing awards for usage of scientific platforms.	• Budget TBD - Includes $200K of AWS credits • To be released no early than April

A NASA OPEN-SOURCE SCIENCE INITIATIVE:
TOPS: TRANSFORM TO OPEN SCIENCE

Leading the Path to Open-Source Science

Transform to Open Science (TOPS) is a $40 million* 5-year NASA Science Mission Directorate mission

Objectives:

- ★ Increase understanding & adoption of open science.
- ★ Accelerate major scientific discoveries.
- ★ Broaden participation by historically underrepresented communities.

Year of Open Science

Goals for 2027:

- ★ 20K earn Open Science Badge
- ★ 5+ major discoveries
- ★ Increase participation of underrepresented groups by 2x

*pending appropriations

2023 is NASA's Year of Open Science

TOPS will be energizing and uplifting open science across the scientific community through:

Visibility

Capacity Sharing Resources

Incentives

Moving towards openness

TOPS in the News! We are Everywhere!

TOPS and Year of Open Science **Visibility**

- Agency comms
- Articles
- Announcements
- Twitter Spaces
- Community meetings

Conference **Visibility**

Annual 2023 Meeting: Open Science theme
Promote & Launch the TOPS Open Science Course
Booths, Events, Workshops, Plenary Talks, Comms
AGU, AMS, AAS, AAAS, and more…

Capacity Sharing : Resources

- Open Science Course in Open edX
  - High quality, interaction Open Online Course
  - Free, public, open - for in-person, virtual, and independent learners
  - Videos / quiz / interactive activities/workbooks
  - Fast-pass option for experienced open science practitioners
  - Open edX LMS tracks learners, completion of modules, data analytics

- Incentivize completion of course
  - Gamification: Certification / badges
  - Prizes, challenges, and bootcamps

- Make it easy & everywhere
  - Workshops at all big meetings
  - Workshops at science team meetings
  - Workshops through virtual cohorts

Capacity Sharing - Resources: Open Science Curricula
5 Modules Organized as a Scientific Workflow

What is open science, why does it benefit me, and why does it benefit the greater scientific community?

How to share software

Best practices for sharing all results and analysis, as well as peer reviewing

ETHOS OF OPEN SCIENCE | OPEN TOOLS & RESOURCES | OPEN SOFTWARE | OPEN DATA | OPEN RESULTS

How to use popular open science tools

How to effectively use and share open data

Earn Badges at Each Level

Complete All 5 & earn TOPS Open Science Badge & Certification

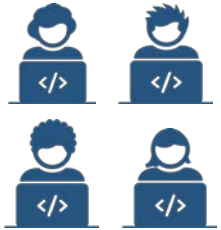
Capacity Sharing within the Community

TOPS Champions

Scientists to help teach modules at events and act as Open Science champions

Cohorts

Engage with learners through a virtual cohort model to increase Open Science Badge achievement



Summer Schools

Institutions selected to run 8-12 weeks of teaching the 5 modules to selected science teams + open competitive student/early career researchers

Curriculum Expansion

Groups funded to migrate/create discipline specific modules and data science skills modules to Open edX TOPS platform

Hackathons

More hackathons that advance data science skills and open science

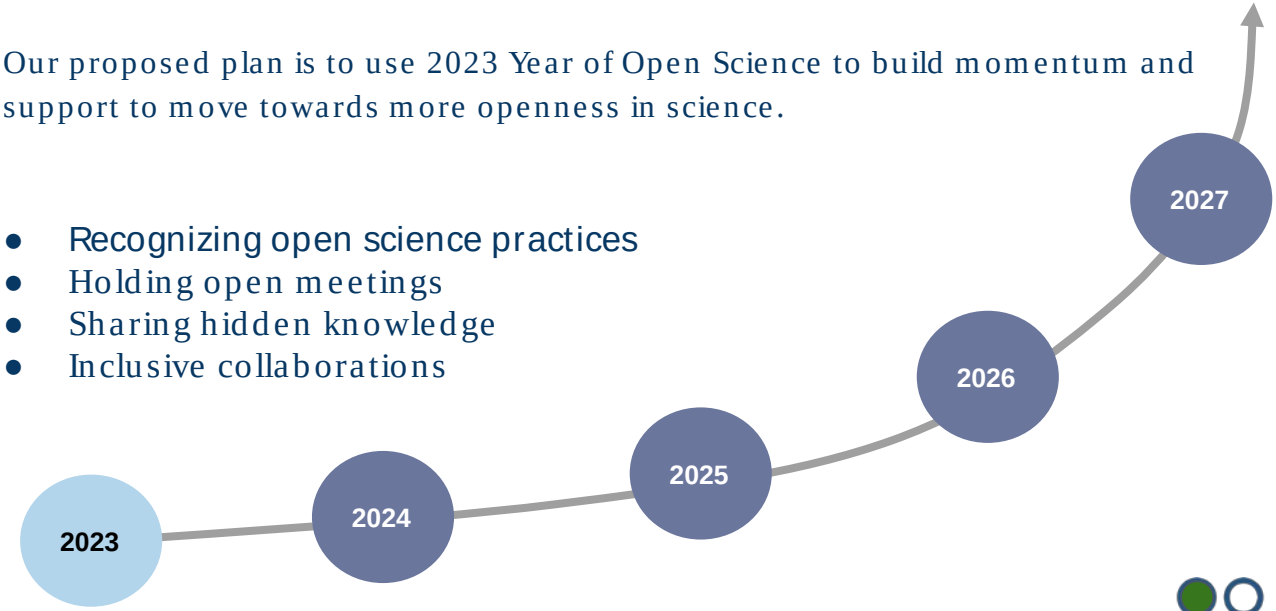
Incentives: Open Science Awards

- Societies create & manage TOPS Open Science Prizes & Awards programs
  - Award Purpose: To reward significant leadership and progress toward open science and showcase the benefits of open science
- Work with societies to evaluate and update their existing awards and recognitions to:
  - Include open science activities as review criteria
  - Where possible allow for team nominations

Moving towards openness: Year of Open Science and the Future

Our proposed plan is to use 2023 Year of Open Science to build momentum and support to move towards more openness in science.

- Recognizing open science practices
- Holding open meetings
- Sharing hidden knowledge
- Inclusive collaborations



Update necessary systems to increase visibility

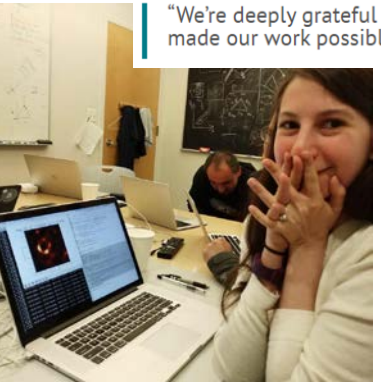
Require a little

Require a little more

Require a lot

Open Science Results Speak for Themselves...

"We're deeply grateful to all the open source contributors who made our work possible." –Dr. Katie Bouman



First image of black hole

"The open source community is very important for scientists; imagine if we had to do everything from scratch every single time." –Dr. Chi-Kwan Chan

We "greatly improve[d] our own work by adopting well-tested community packages that contain the collected wisdom of many other projects." –Dr. Lindy Blackburn

"with the open source projects in NumFOCUS, we were able to iterate our algorithms so fast that they enabled us to finish our work in two years"

Dr. Julia Stewart Lowndes @juliesquid

Replying to @ChelleGentemann and @theNASEM

Congrats Chelle!
The welcoming, inclusive, collaborate-and-reuse culture of the #rstats community is something that changed my science-life and my life-life. Hard to distill but here are a few attempts:
openscapes.org/blog/2020/02/2...
openscapes.org/blog/2019/02/1...
openscapes.org/blog/2019/08/2...

3:15 PM · Mar 11, 2022 · Twitter Web App

Lucas Sterzinger @lucassterzinger

Replying to @ChelleGentemann and @theNASEM

Probably the most common answer, but using @xarray_dev, @dask_dev, @ProjectJupyter, and @matplotlib has been the backbone of my research since day 1. Working with these tools also motivates me to make the data and code for my plots open source, making my science more reproducible

7:41 AM · Mar 11, 2022 · Twitter Web App

Pierre de Buyl @pdebuyl

Replying to @ChelleGentemann and @theNASEM

In remote sensing: using @PyTrollOrg satpy as a comparison point for reading geostationary satellite data, @scitools_iris and panoply from @NASA for plotting said data.

12:15 PM · Mar 11, 2022 · Twitter Web App

Sam Ehrenstein @elasticsnake

Replying to @ChelleGentemann and @theNASEM

In computer science, research moves very fast. It would not be possible to keep up with the latest work if not for the arXiv and open-access conferences.

1:47 PM · Mar 14, 2022 · Twitter W

Ricardo Barros Lourenço @rblourenco

Replying to @ChelleGentemann and @theNASEM

I've briefly returned to the public-private sector (between 2019-21) and the nicest thing about working with OSS during all my career was the ability to show new methods to be applied in that company, which was of clear understanding, helping auditing efforts.

7:56 AM · Mar 12, 2022 · Twitter Web App

Scott Collis (He/Him) @Cyclogenesis_au

Replying to @ChelleGentemann @openscience and @theNASEM

Being an open scientist has:
1) accelerated my career. It has allowed me to choose projects which benefit more people.
2) Has created long lasting collaborations and friendships. When you are open you are... open!
3) Made me a better scientist. "Show your working!"

6:36 AM · Mar 12, 2022 · Twitter Web App

Paola Masuzzo @pcmasuzzo

Replying to @ChelleGentemann and @theNASEM

An aspect we should talk more about, open research practices as a driver to a real reform in the research endeavour. I try to depict it in this image :)

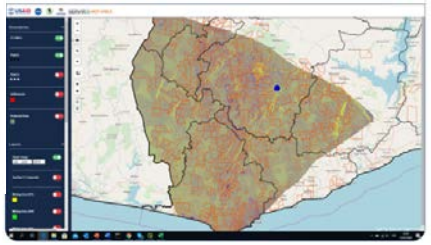
Belize GEO @BzGEO · Mar 11

Replying to @ChelleGentemann and @theNASEM

Our friends @SERVIRGlobal have many examples of how algorithms + code from one region have been customized for use in another. An example is gold mining monitoring, where Amazonia + W. Africa have collaborated in an #OpenScience context, leveraging #GEE.

simonestaiger @simonestaiger · Apr 8, 2020

Reducing illegal gold mining in the tropical forests of Ghana and Peru: A forthcoming collaboration across the Atlantic #SERVIRamazonia servir.ciat.cgiar.org/illegal-gold-m... @USAIDPeru @SERVIRGlobal @CERSGIS_GH @NovoaSidney @amazonacca @sig_gis @BiovIntCIAT_eng



Max Grover @mgroverwx · Mar 11

Replying to @ChelleGentemann and @theNASEM

Here's a great use-case of @Py_ART , which is funded by @doescience @armnewsteam ! Over 200 citations so far, with many including awesome code like this paper which enables #OpenScience !

Milind Sharma @Gewitter_Blitz · Mar 11

The power of open source software! The authors (@jehcssou and @deeplycloudy) also provide a clean code to encourage reproducible science. I could apply their technique to my dataset within a few hours. Neat! Yes to #OpenScience

Quotes from https://numfocus.org/case-studies/first-photograph-black-hole


Questions?