Your Mission's Open Science Journey

A "Choose Your Own Adventure"

Rebecca Ringuette
on behalf of many contributors
and supporters

Link to this slide deck:
https://docs.google.com/presentation/d/1dSlMZwVBZwCVabzK4nulz_C95PhjOXA597M8KavHidw/edit?usp=sharing

Organizational Notes

- This version incorporates the comments entered before Oct 3, 2023, and has been converted to a poster to be presented at DASH http://dash.heliophysics.net/
- We will meet after DASH to incorporate additional feedback (weeks of Oct 23 or Nov 7). Please email Rebecca Ringuette to be added to the email list for these notifications (email on next slide).
- **If you would like to present these slides elsewhere or otherwise be involved, please comment here to let us know!**
- We are working towards a 2-day workshop in 2024 to draft a rubric. Organizers/hosts are welcome.

Participation Instructions

- **Please add comments** to these slides to join the conversation, and add your name to the next slide for contributions or the following one for support. Please be mindful that all comments on these slides are **public**. (This slide is repeated later in slide 18.)
- Look over the following slides and **choose 1-3 of your favorite tasks** described.
- If you have an idea for an open science task that is not in the slides, please comment on the last slide of tasks with the idea.
- **Leave a comment** on the related slide with
  - The mission, institution, or project you are associated with,
  - Your chosen favorites,
  - Why you chose these, and
  - What benefit you expect to or have already seen from these tasks.
- Alternatively, you can **comment on the task** with this feedback.
- Your comments will be added to future versions of these slides and **openly shared** with the community.

We look forward to your input!

Contributors

General Contributors (not associated with a mission):

- Rebecca Ringuette / Heliophysics Digital Resource Library
  (Point of contact for this work: rebecca.ringuette[at]nasa.gov)
- Julie Barnum / The Python in Heliophysics Community
- Caroline Coward, Brian Knosp, Daniel Limonadi / Jet Propulsion Laboratory
- Rachel Paseka / NASA HQ

Missions (please add names/institutions here):

- PUNCH Mission: Barbara Thompson/GSFC, Matthew West/SwRI, Marcus Hughes/SwRI, Dan Seaton/SwRI, Chris Lowder/SwRI, Ritesh Patel/SwRI, Craig DeForest/SwRI
- GDC Mission: Doug Rowland/GSFC, Sarah Dietrich/GSFC, Amy Rager/GSFC
- HelioSwarm Mission: Jonathan Niehof/UNH
- STEREO & SDO: Barbara Thompson/GSFC

Supporters

This is where we list those who support the ideas presented in these slides but have not contributed (similar to co-signers on a white paper). Just add a comment in the appropriate category with your name and institution(s). Happy to include you!

Missions (please add missions: names/institutions here):

General Supporters (please add names/institutions here):

- Lan Jian, Brian Thomas / Heliophysics Digital Resource Library
- Natasha Batalha / NASA
- Leah Wasser / pyOpenSci
- Jenn Gannon, Sudha Kapali / CPI
- Viv Hutchison / USGS
- Ryan M. McGranaghan / Jet Propulsion Laboratory
- Jon Vandegriff / JHU/APL
- Greg Lucas / LASP

Your Mission's Open Science Journey

Goals - Why should a mission do this?

- Give missions the opportunity to **decide how they want to be judged** in their open science activities.
- Help missions become **more competitive** in their proposals.
- Create a welcoming, open, collaborative **atmosphere**.
- **Advance mission science** through open collaboration.
- Improve data products and software **through collaborative efforts**.
- **Minimize duplication** of efforts by reusing and a
- **Build public trust** in and ownership of mission results through openly sharing work in progress and community engagement.
- **Guard against negative influences** through DOIs and collaboration and conduct rules.
- **Tell the story** of the mission to the wider community to increase future funding.

Image Credit: https://science.nasa.gov/science-red/s3fs-public/atoms/files/SMD%20Open-Source%20Science%20Guidance%20Version%202.0_0.pdf

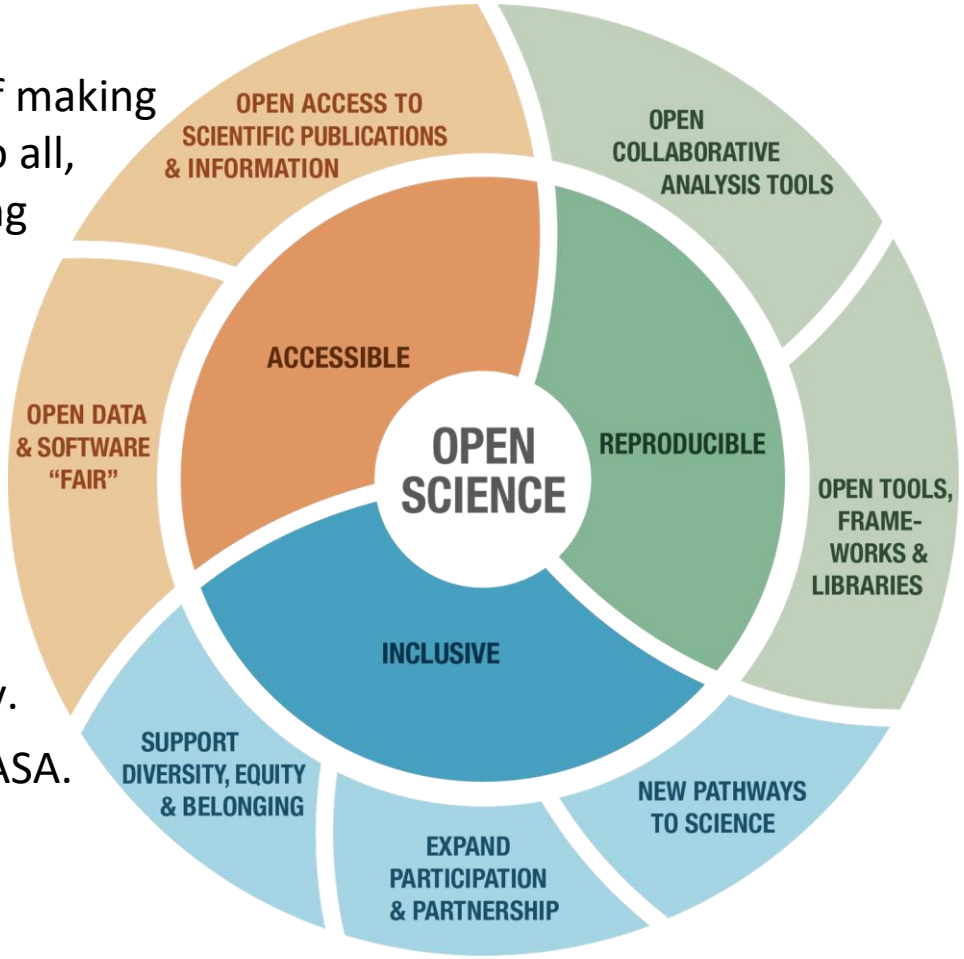
Open Science Overview

What is Open Science?

Open Science is the **principle and practice** of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility and equity.

Why open science?

- Accelerates scientific discovery.
- Greater collaboration and efficiency.
- Enhanced transparency and reproducibility.
- Mandated by the U.S. White House and NASA.



Definition source:
https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research/

Image Credit: NASA TOPS
https://zenodo.org/record/6565080#.ZFPvCnbMKUk

Open Science Overview

In January 2023, the White House, 10 Federal Agencies, and a coalition of universities and organizations declared 2023 to be a "**Year of Open Science**."

Now, many organizations and 16 federal agencies have open science initiatives:

https://open.science.gov/

The Year of Open Science has 4 goals:

1. Establish strategic approaches for advancing open science,
2. Promote equitable participation in open science through transparency, integrity and equity of reviews,
3. Account for open science activities in evaluations and incentives, and
4. Engage underrepresented communities in the advancement of open science and research.

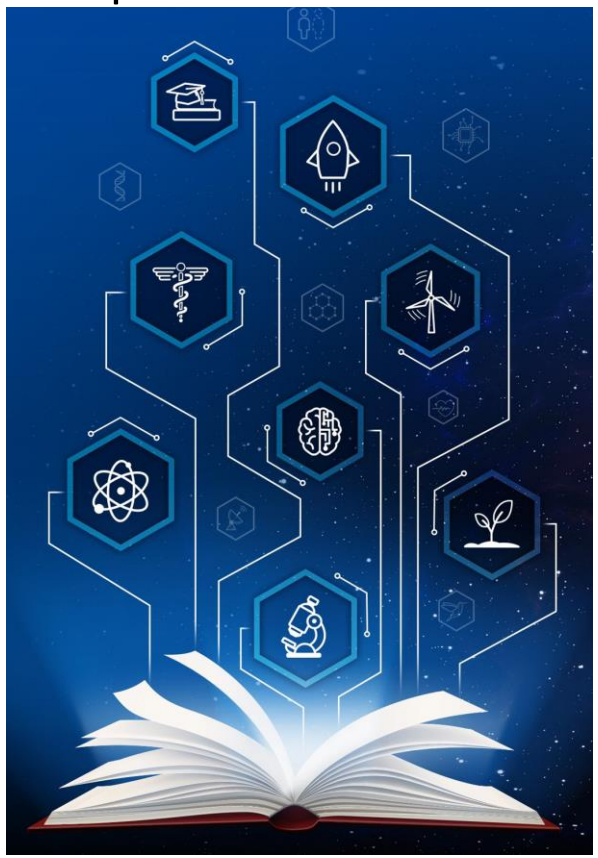


Image Credit: https://nasa.github.io/Transform-to-Open-Science/year-of-open-science/

Open Science Overview

NASA's commitment to open science over the next decade is embodied in the **Open-Source Science Initiative (OSSI).**
NASA's **Transform to Open Science (TOPS)** is the community engagement component of OSSI.

OPEN (**TRANSPARENT**) SCIENCE
scientific process and results should be visible, accessible, and understandable

OPEN **(ACCESSIBLE)** SCIENCE
data, tools, software, documentation, and publications should be accessible to all (FAIR)

OPEN **(INCLUSIVE)** SCIENCE
process and participants should welcome participation by and collaboration with diverse people and organizations

OPEN (**REPRODUCIBLE**) SCIENCE
scientific process and results should be open such that they are reproducible by members of the community

Open Science Overview

The TOPS initiative recommends that **individuals** complete five training modules (Open Science 101):

- Ethos of Open Science
- Open Tools and Resources
- Open Data
- Open Software
- Open Results

There are additional **recommendations** for teams, organizations and projects. TOPS has three supplemental open science guides: individuals, teams, and organizations (see links at end).

NASA's SPD-41a provides a set of requirements for publications, data, software and other scientific knowledge and specifically apply to NASA-funded missions. **Each division of NASA SMD has additional requirements**, typically in the form of calibration plans, data and software management plans, and similar documents.

We can build upon these requirements and open science principles to create guidelines specific to an open mission!

Open Science Overview

Open Science is not a binary choice.

- *It is a spectrum of possibilities.* Beyond the minimum requirements, missions (and researchers) can choose how far up the spectrum they want to be for each component.
- Missions are often limited by legal restrictions, but can choose other ways to advance.

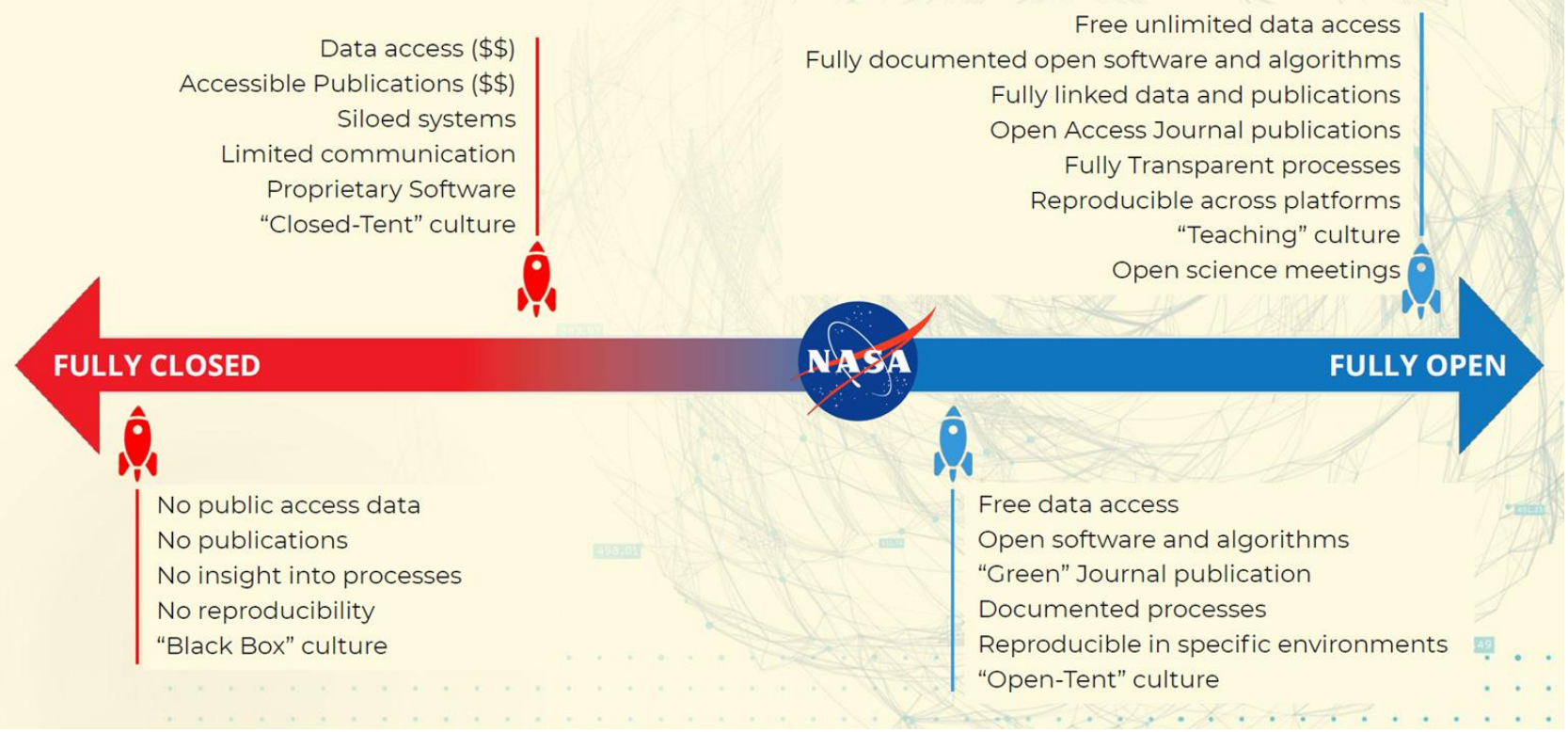


Image Credit: NASA TOPS, https://doi.org/10.5281/zenodo.5621673, slide 19.

Your Mission's Open Science Journey

How does a mission incorporate open science principles?

- **Infrastructure is building** processes to support SPD-41a requirements.
- Begin **planning for open** publications, data, and software by navigating legal issues now (e.g. licensing).
- **Find tools** to make your processes open and easy (e.g. GitHub and PyHC).
- Choose **additional practices** to add value and open the mission's culture.

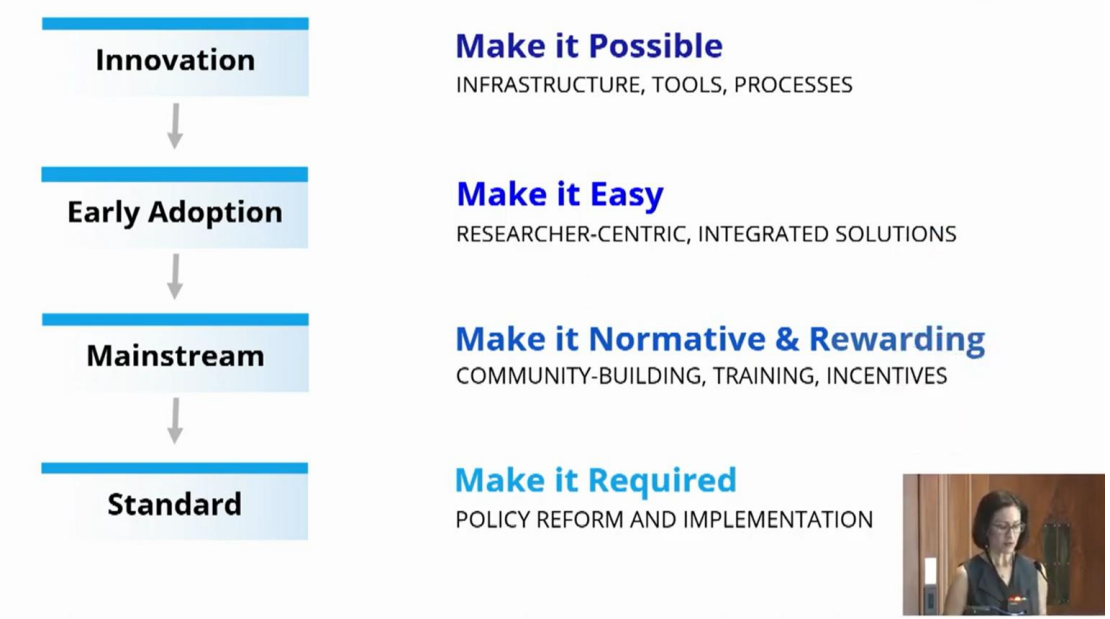


Image credit: Lisa Cuevas Shaw, Chief Operating Officer and Managing Director, Center for Open Science. The session was part of the COS 10 Year Anniversary, which took place in Washington, DC on May 8, 2023.
https://www.youtube.com/watch?v=dgB8Ry_cRwE&list=PLChfyH8TVDGmWyGePxToZ4ZcNiBkFKubI&index=5

SPD-41a: The Starting Line

SPD-41a applies to:

- All NASA-funded missions **pre-KDPb as of Mar 2, 2023 are required** to satisfy all of the 'SHALL' (or must) statements in SPD-41a. 'SHOULD' can be interpreted as 'please'.
- NASA-funded missions in **later stages are encouraged** to select tasks that are feasible with current funding.
- Applies to publications, data and software (see policy for examples):
  - **Publications**: Scientific and technical documents released through print, electronic, or alternative media.
  - **Data**: Scientific or technically relevant information that can be stored digitally and accessed electronically.
  - **Software**: Computer programs, including source and object code, that provide users some degree of utility or service.
  - **Restrictions** are acknowledged and allowed for. See policy for details.

Where to report these tasks and any restrictions?

- Your Open Science and Data Management Plan (OSDMP - see links at the end for templates).
- Your proposal text.

SPD-41a: The Starting Line

Starting with current requirements: SPD-41a

(see policy for full details, links on last slide. Yellow = SHALL/MUST, none = SHOULD/Please)

- *Publications*:
  - Peer-reviewed publications **immediately** available on NASA's pubspace,
  - Content from mission workshops available publicly.
- *Data*:
  - **Open and publicly available** as soon as possible, without fees or restrictions, and including calibrations and similar items
  - **Findable** and machine-readable FAIR data and metadata **aligned** with community standards,
  - Produced in a open **standard format** common in the community (e.g. CDF, netCDF, FITS for Heliophysics),
  - Have an **open license** (preferably CC0) and a persistent identifier (DOI).
- *Software*:
  - Released and **developed in the open** when restrictions allow,
  - **Reported** to NASA for indexing and available in a public repository,
  - Have a persistent identifier (DOI),
  - Follow best practices, and
  - Have an **open license**, code of conduct and contribution guidelines.

SPD-41a: The Starting Line

A "should/please" in SPD-41a

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Image Credit: Wilkinson *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Your Mission's Open Science Journey

- SPD-41a defines the **minimum requirement** for NASA-funded missions.
- Missions have trouble meeting these requirements due to restrictions.
- Each mission can **choose activities beyond the minimum** requirements to:
  - ***Make its OSDMP more competitive in the selection process.***
  - enhance its culture,
  - advance the mission science,
  - supplement its openness when restrictions prevent in other areas, and
  - build public trust in and ownership of the mission results.
- The lists of additional open science tasks on the following slides are categorized into **pre-proposal, pre-launch, and post-launch** categories, with further divisions into *culture* and *artifact* related tasks.
- Missions can choose to **complete tasks related to any stage** regardless of their current status.
- The lists of tasks in later categories do not repeat the tasks of earlier categories for the sake of brevity, but the earlier tasks are still relevant.

Your Mission's Open Science Journey

- These slides
  - summarized the concepts of Open Science,
  - outlined NASA's requirements for missions, and
  - present (in the slides at the end) detailed suggested activities at various stages relevant to missions.
- ***Open Science is not a binary choice.*** There is a spectrum of possibilities for each aspect of the mission.
- Supplementing a mission's OSDMP with additional tasks help make a mission **more competitive**.
- See the links at the end of this presentation for more details, including the **TOPS training modules on open science**.

Your mission's open science journey is an adventure YOU choose.

Participation Instructions

- Look over the following slides and **choose 1-3 of your favorite tasks** described.
- If you have an idea for an open science task that is not in the slides, please comment on the last slide of tasks with the idea.
- **Leave a comment** on this slide with
  - The mission, institution, or project you are associated with,
  - Your chosen favorites,
  - Why you chose these, and
  - What benefit you expect to or have already seen from these tasks.
- Alternatively, you can **comment on the task** with this feedback.
- Your comments will be added to future versions of these slides and **openly shared** with the community (beginning on the next slide).

We look forward to your input!

Community Feedback

- ***Feedback from the community starts on this slide, followed by the list of possible tasks missions can choose from. The related task is repeated here, followed by the signed comment in a smaller font.***
- Mission artifacts

  This is where I put most of my votes. Making the mission artifacts accessible now in the way they will be accessible later is crucial to validating them and equalizing the playing field for everyone. Missions often keep important data at instrument sites, and that does not survive at a generic archive, so having missions rely on the archive during the mission could force things to all go into the archive. - Jon Vandegriff, JHU/APL
- Collaboratively developed software to increase public ownership

  This can benefit missions if the software has a life of it's own and lives beyond the mission; this is currently not true of mission-specific portals, but we could make open source portal frameworks (this is hard).
  - Jon Vandegriff, JHU/APL
- Collaboratively developed software to increase public ownership … Reviewed by pyOpenSci, JOSS, or the appropriate review process.

  Any python packages that are in scope for JOSS can be reviewed by us and accepted by JOSS too. We also offer a pyopensci approved badge for the readme / software repo, AND for those who want to be involved on the software side, we are always looking for reviewers and building out our editor team. - Leah Wasser, pyOpenSci
- Mission artifacts that enable Open Data produced by real-time systems.

  For example: serve data through Rest API in real-time, transfer data to public repositories, follow FAIR principles in data validation and management. This is especially relevant for automated solutions for validating data and metadata tagging in real-time. - Sudha Kapali, CPI

Choosing Your Own Adventure: Pre-Proposal

Mission Culture

- Team leaders and members **become familiar** with the practices of open science (and complete the TOPS training modules when available).
- **Discuss** ways to incorporate open science practices in all mission aspects.
- **Draft an open science policy** for your mission.
- **Appoint** team members to be responsible for guiding open science practices.
- Recognize open science activities in **hiring practices**.
- Construct a code of conduct and collaboration rules.
- Define a **clear way for potential collaborators** to contact mission and science working group leaders.
- Define what activities will be **open to non team members** and how to track their contributions.
- Sign the DORA statement (https://sfdora.org/)

Choosing Your Own Adventure: Pre-Proposal

Mission Artifacts

- If allowed, **post research plans** (with DOIs!) on the OSF project pages.
- **Share sample datasets** in a public repository with a DOI.
  (preferably SDAC or SPDF or other domain-specific archive)
- Find and **collaborate with open source software** to plan out algorithm needs.
- Work through possible **legal limitations**, such as licensing and permissions for research plans, datasets, and software to be able to develop and share openly.
- Plan how to **preserve and share** slide decks, posters, and similar material from all meetings from a single publicly accessible location (e.g. Zenodo collection or OSF Meetings).

Include all completed items **in your proposal's OSDMP** or similar document.

Choosing Your Own Adventure: Pre-Launch

Mission Culture

- Does the mission provide **information** specifically geared towards scientists who are not on the team (non-team-members)?
- Are there **events and activities** that promote collaboration with the community (open science meetings, working groups, special sessions)?
- **Communicate** to the public through open access platforms (like HelioNauts, GitHub discussions, and Slack) to ensure everyone has access to knowledge and solutions.
- Add **short routine updates** (e.g. every 1-2 months) on each group's page on what the group is working on or has recently achieved (e.g. 1-2 paragraphs with a graphic or group picture).
- Make your **website easily findable** by including appropriate tags for internet search engines to catch (e.g. space weather, heliophysics, your science topic).
- Create and use an **open science points system** for individuals and teams to be acknowledged for their efforts and earn rewards.
- Determine authorship order based on a **public standardized rubric for contributions** to publications, data and software, and including non-team members' contributions.

Choosing Your Own Adventure: Pre-Launch

Mission Artifacts

- Create a network of **public projects with DOIs** (e.g. one subproject per working group) on the Open Science Framework (https://osf.io/).
- Create and update **research plans** (preregistrations) on OSF to document decisions in development and analysis methods.
- **Share updated sample datasets** in a public repository with a DOI.
- **Collaborate** with existing open source software to accelerate development.
- Create a **public code repository** with a DOI for each section of software.
- **Collaborate with the public** on documentation and tests for those softwares.
- Use best practices for software development (see https://www.pyopensci.org)
- Create a book of **executable notebooks** demonstrating various portions of the data processing, calibrations, and science analyses.
- **Preserve and share** slide decks, posters, and similar material from all meetings from a single publicly accessible location
  (e.g. Zenodo collection, OSF Meetings, or the mission's website).
- Create a **RestAPI to serve data** through (in real-time if possible).

Choosing Your Own Adventure: Post-Launch

Mission Culture

- Continue to use open platforms to c**ommunicate progress** to the community.
- Add **short routine updates** (e.g. every 1-2 months) on each groups page on what the group is working on or has recently achieved (e.g. 1-2 paragraphs with a graphic or group picture).
- Discuss and create **citizen science opportunities** using mission data and artifacts.
- **Contribute a success story** to the NASA TOPS effort.
- **Provide feedback** to the NASA TOPS open science recommendations for missions.
- Use the TOPS and OSF badges to **recognize significant open science accomplishments** of each science working group (see https://osf.io/tvyxz/ for examples).
- Make these badges public on the website and related pages (e.g. code repositories, OSF project pages/subpages, etc). **Wear them proudly!**
- Routinely **recognize leadership in open science activities** in mission-wide emails, but also include other forms of leadership and initiative.
- **Present** on these efforts at major meetings and venues each year.

Choosing Your Own Adventure: Post-Launch

Mission Artifacts

- Serve data through a RestAPI and **through domain-specific repositories** (in real-time if possible).
- Obtain a **CC0 license for the data** and an **open license for the software** (as possible).
- Develop and iterate on calibration and data processing software openly to advance mission science through **collaborations with the community**.
- Update and **collaborate on executable notebooks** demonstrating various portions of the data processing, calibrations, and science analyses.
- Build and transfer mission software **into an existing open source software packages** for long-term availability and maintenance.
- Create a new DOI for each software version associated with a data version, and **connect the DOIs** for the software and data versions in the metadata.
- **Create software documentation and tutorials**, preferably interactive.
- Continue to use best practices in software development and coding styles.
- **Transfer** mission publications, data, and software **to domain-specific archives**.
- Conduct routine (e.g. yearly) **competitions** between science working groups to **reproduce the other groups' work**. Include the public on this via hackathons. *(This work can be published!)*

Choosing Your Own Adventure: Post-Launch

- ***Validatable publications to build public trust***
  - Archive and describe all related items needed for validation with the domain-specific archive (e.g. intermediate and final datasets, software, instructions), preferably in a hosted cloud environment,
  - Link to relevant version-controlled preregistrations,
  - Link to previous reproduction efforts, and
  - Link validation items to the related publication via metadata curation.
- ***Transparent data to build public confidence***
  - Processed with a software open for public review and contribution,
  - Available in multiple-processing levels with all items needed for validation, preferably on the cloud with related software while the mission is active or recent,
  - Include re-use instructions with interactive examples.
- ***Collaboratively developed software to increase public ownership***
  - Develop processing and access software in the open, with contact information,
  - Build on existing open source packages (e.g. PyHC packages) and in free languages,
  - Described with standard metadata, with descriptive and interactive documentation,
  - Included and searchable in a domain-specific archive,
  - Reviewed by pyOpenSci, JOSS, or the appropriate review process, and
  - Findable in a domain-specific search interface and by its DOI.

Choosing Your Own Adventure: Additional Ideas

- Use **recognition vocabularies** to recognize contributions to data, software, and publications (see links on last slide).
- Choose to provide **higher quality metadata** (e.g. better descriptions and keywords) when archiving items.
- Add your ideas here!

References

NASA Open Science references

- Year of Open Science
  https://nasa.github.io/Transform-to-Open-Science/year-of-open-science/
- Open Source Science Initiative (including TOPS training modules!)
  https://science.nasa.gov/open-science-overview
- Open Science for Teams guide: https://nasa.github.io/Transform-to-Open-Science-Book/Open_Science_Cookbook/Your_Teams_Open_Science_Journey.html
- Open Science for Organizations guide: https://nasa.github.io/Transform-to-Open-Science-Book/Open_Science_Cookbook/Your_Organizations_Open_Science_Journey.html
- Open Science Success Stories:
  https://science.nasa.gov/open-science/transform-to-open-science/stories/
- NASA SMD Open Source Science Guidance:
  https://github.com/nasa/smd-open-science-guidelines/tree/main
- NASA OSS Guidance on OSDMPs:
  https://github.com/nasa/smd-open-science-guidelines/blob/main/OSS_Guidance/OSDMP.md
- Sign up for NASA TOPS Open Science 101:
  https://nasa.github.io/Transform-to-Open-Science/signup/

References

NASA Policy Links

- **NASA's SPD 41a: https://smd-cms.nasa.gov/wp-content/uploads/2023/08/smd-information-policy-spd-41a.pdf**
- **SMD Science Information Policy page, especially the FAQ: https://science.nasa.gov/researchers/open-science/science-information-policy/**
- **NASA's Open Source Science Guidance (especially p. 5 for division-specific links): https://science.nasa.gov/science-red/s3fs-public/atoms/files/SMD%20Open-Source%20Science%20Guidance%20v2%2020230407.pdf**
- **Compilation of NASA SMD Division policies: https://github.com/nasa/smd-open-science-guidelines/blob/main/OSS_Guidance/SMD_Division_Policies.md**
- **NASA's Public Access Plan (under revision): https://www.nasa.gov/sites/default/files/atoms/files/nasa_ocs_public_access_plan_may_2023.pdf**
- **NASA's NPR 2210: https://nodis3.gsfc.nasa.gov/npg_img/N_PR_2210_001E_/N_PR_2210_001E_.pdf**
- **NASA's PubSpace: https://ntrs.nasa.gov/collections/pubspace**

References

General Open Science references

- Van Tuyl, Steve (Ed.). (2023). Hiring, Managing, and Retaining Data Scientists and Research Software Engineers in Academia: A Career Guidebook from ADSA and US-RSE. Zenodo. https://doi.org/10.5281/zenodo.8274378
- "Introducing Open Science and the OSF" by Gretchen Gueguen: https://docs.google.com/presentation/d/1vtSmbsDweTLmS8wGgfwNu9aDQMZ3iMnPQeQTlKJje_M/edit#slide=id.g23f52b916f5_0_0
- The Center for Open Science 10th Anniversary meeting: https://www.youtube.com/playlist?list=PLChfyH8TVDGmWyGePxToZ4ZcNiBkFKubI
- The Open Science Framework: https://osf.io/
- The Declaration on Research Assessment (DORA): https://sfdora.org/
- pyOpenSci: https://www.pyopensci.org/
- Python in Heliophysics Community: https://pyhc.org/

Know of a useful link? Add it in a comment on these slides!

References

Standardizing Recognition

- **Contributions to publications:**
  https://credit.niso.org/
- **Contributions to datasets (e.g.):**
  Beirer et al. (2017). Data Authorship as an Incentive to Data Sharing.
  https://doi.org/10.1056/NEJMsb1616595
- **Contributions to software (e.g.):**
  Alliez, P. et al. (2019). Attributing and Referencing (Research) Software: Best Practices and Outlook from Inria. Co. https://doi.org/10.48550/arXiv.1905.11123
  https://allcontributors.org/