Jet Propulsion Laboratory
California Institute of Technology

SPD-41a and you:

Key points for SMD-funded researchers and mission scientists from NASA's new open science policy

Marshall J. Styczinski*
NASA Postdoctoral Program Fellow,
Jet Propulsion Laboratory, California Institute of Technology
marshall.j.styczinski@jpl.nasa.gov

* With suggestions from S. Crawford, SMD Science Data Officer

Disclaimer

- The following is my interpretation of a policy document. I am a scientist and open-source software developer, not a lawyer.
- This is not legal advice.
- This does not represent the position of JPL or NASA and does not constitute any promise on behalf of any entity.
- More resources are available than what I can include here! Some official FAQs and info pages are at the end (31/30).
- **Further questions: send to** HQ-SMD-SPD41@list.nasa.gov

Outline

1. Background — what and why
2. Key points of the new policy
3. Details for mission data/software
4. Details for research data/software
5. Suggestions for open science

What is SPD-41a?

- Science Policy Document 41, revision a
- SPD-41a is the second version of the policy. The first version (SPD-41) was released Aug 2021
- Describes requirements for certain NASA-funded science and missions
- Read at https://science.nasa.gov/science-red/s3fs-public/atoms/files/SMD-information-policy-SPD-41a.pdf

Background

- SPD-41 described existing requirements that applied to the Science Mission Directorate (SMD)
  - Meant to standardize practices when adopted as policy
- SPD-41a requires transparency and access to data and software, in support of "maximizing the benefit of NASA to society"

Who is required to follow SPD-41a?

- All new SMD-funded scientific activities
  - Mission scientific information, for missions not yet in Phase B by 3/2/23
  - Research funded in any way by SMD, including ROSES as of ROSES-23 (does not include ROSES-22)
- Existing projects are encouraged to follow

Why is NASA doing this to us?

- Open science is robust science
  - We all benefit when scientific results are demonstrably reproducible
  - Open access to data and software is required for building on each others' work
  - Mission data and software has max. value when its production is traceable
- Scientific progress is accelerated when we're not spending time reinventing the wheel

Key points of SPD-41a — applicability

For all SMD-funded projects:

- All publications, and all data and software needed to reproduce results/figures, must be shared publicly
  - If developed using SMD funding
  - Except when blocked by ITAR, HIPAA, etc. Those are "restricted software"
  - Does NOT apply to physical objects (lab specimens)
- All unrestricted mission software, including for operations, planning, data acquisition, processing, and analysis

Key points of SPD-41a — data sharing requirements 1

- Data must be archived:
  - In non-proprietary, open formats (e.g. ASCII text, HDF5, CDF)
  - With metadata that describe the data
  - With an open data license, preferably CC0, but other open licenses may be appropriate depending on circumstances
  - With a persistent identifier (DOI)
- Data collections must be listed as “part of the NASA catalog of data.” Being discoverable in the Science Discovery Engine (beta) will likely satisfy the requirement: https://science.data.nasa.gov/

Key points of SPD-41a — data sharing requirements 2

- Data must be archived:
  - If collected by an SMD-funded mission
  - If needed to validate the results of publications, including data shown in figures, maps, and tables
  - If otherwise “scientifically useful”
  - and cited in publications that rely on it
- Data should follow FAIR principles: to be Findable, Accessible, Interoperable, and Reusable

Key points of SPD-41a — software requirements 1

- Software must be made publicly available:
  - With a persistent identifier (DOI) in an archive repository with long-term curation (e.g. Zenodo)
  - Regardless of the language, including proprietary formats
  - If it “provides users some degree of scientific utility or produces a scientific result or service” but not preliminary results. If some software is needed to produce the figures in your publication, it needs to be shared.
  - With a code of conduct and guidelines for contribution
  - and indexed “as part of the NASA catalog of software” at: https://invention.nasa.gov/ (single-use excepted from this)

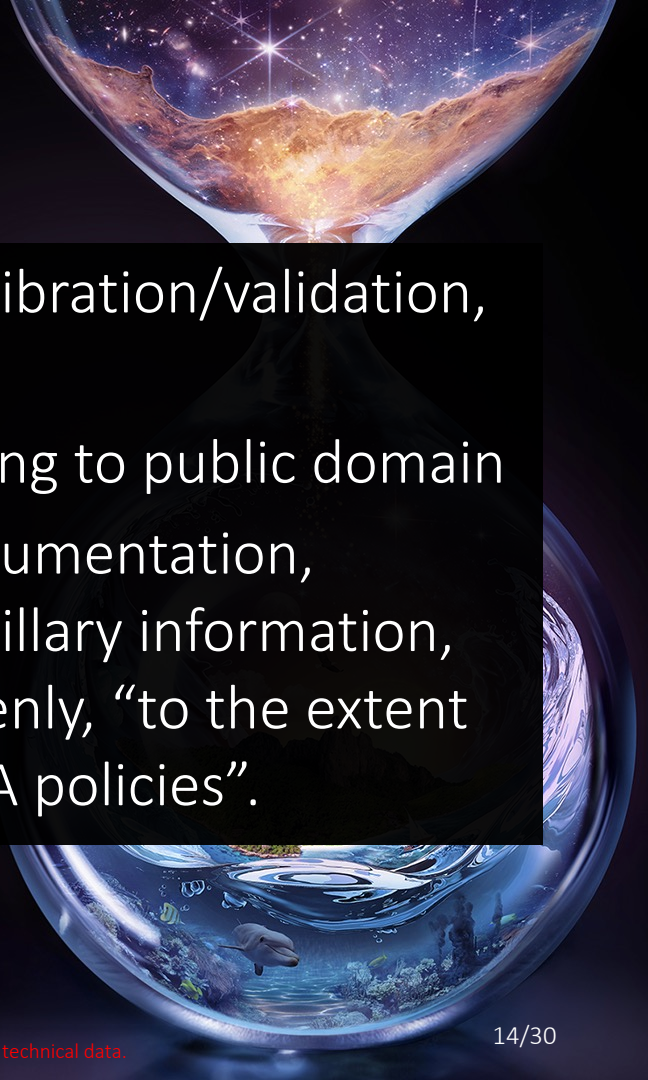
Key points of SPD-41a — software requirements 2

- Software should be made publicly available:
  - If enhancements (new functionality) were made to an existing project that was not previously openly available
- Shared software should:
  - Follow best practices, such as style guides for code readability, comments to explain how it works, have a README file explaining how to install, etc. Your code is more useful than you think! Write it with other scientists in mind.
  - Be cited in publications that use it

Key points of SPD-41a — publications

- As-accepted manuscript versions **must** be deposited in a NASA-approved repository (e.g. PubSpace) as of the publication date, with no embargo period.
  - This was already required by ROSES but with a 1-year grace period after publication
  - As-accepted drafts meet the requirement (“green” open access)
  - JPL personnel: This is done automatically through the Unlimited Release System (URS). Once you submit your accepted draft to URS, it is posted for you.

Details for mission data



- Should be shared immediately following calibration/validation, <6 months after obtained
- Should be shared with CC0 license: ~donating to public domain
- Observations, calibrations, coefficients, documentation, algorithms, software, technical reports, ancillary information, and work products should all be shared openly, “to the extent allowed by applicable law and existing NASA policies”.

Details for mission software

- Must be “developed openly in a publicly accessible, version-controlled platform that allows for contributions and engagement from the community”
  - “Contributions … from the community” = pull requests, issues, etc. GitHub and similar services already include these.
    - You are NOT required to accept public contributions, just the ability is req
  - Most mission software is likely already in such a platform, but held in private version control systems.
- DOES NOT APPLY to ITAR-restricted, command & control, limited release software, etc.

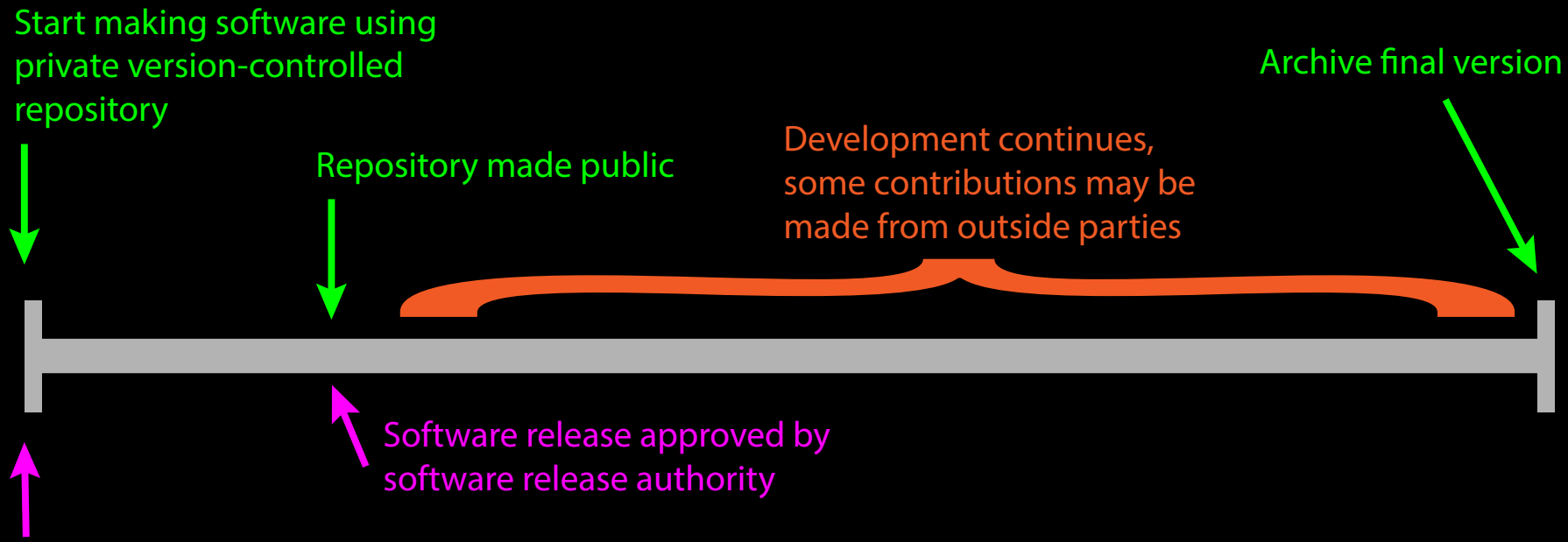
Details for research data

- Must be archived:
  - If scientifically useful, along with any associated publication or the end of the funding period, whichever comes first
- Does not include preliminary analyses, lab notebooks, communications with colleagues, etc.
- Use your judgement: Is it needed to reproduce your results? If so, archive it and cite it.
- Each SMD division provides further guidance

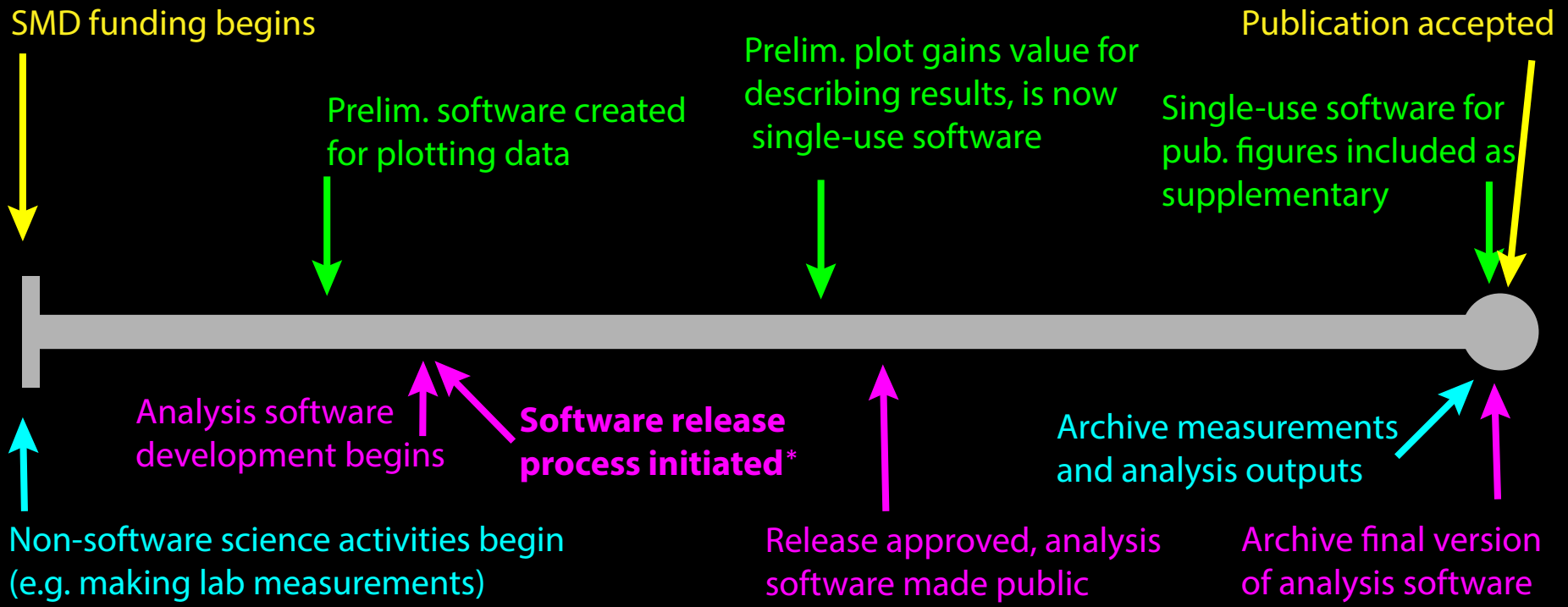
Details for research software

- Must be released:
  - If it provides “some degree of utility or produce[s] a result”
  - As open-source software “no later than the publication date” or the end of the funding period, whichever is first
  - If it pertains to processing of sensor data, models, simulations, or analysis of measurements or models
- Single-use software (e.g. for making a specific plot) can be released as supplementary material or in an archive
- Shared software must be citable with a persistent identifier (DOI). This means versions must also be archived.

Open science under SPD-41a for mission software



Open science under SPD-41a for research software & data



*See JPL Rules! DocID 56592

marshall.j.styczinski@jpl.nasa.gov

SPD-41a is the baseline!

- Each ROSES program element may set specific requirements, such as:
  - Required repositories
  - Required license types
  - Required data formats
- Divisions have provided their own specific requirements

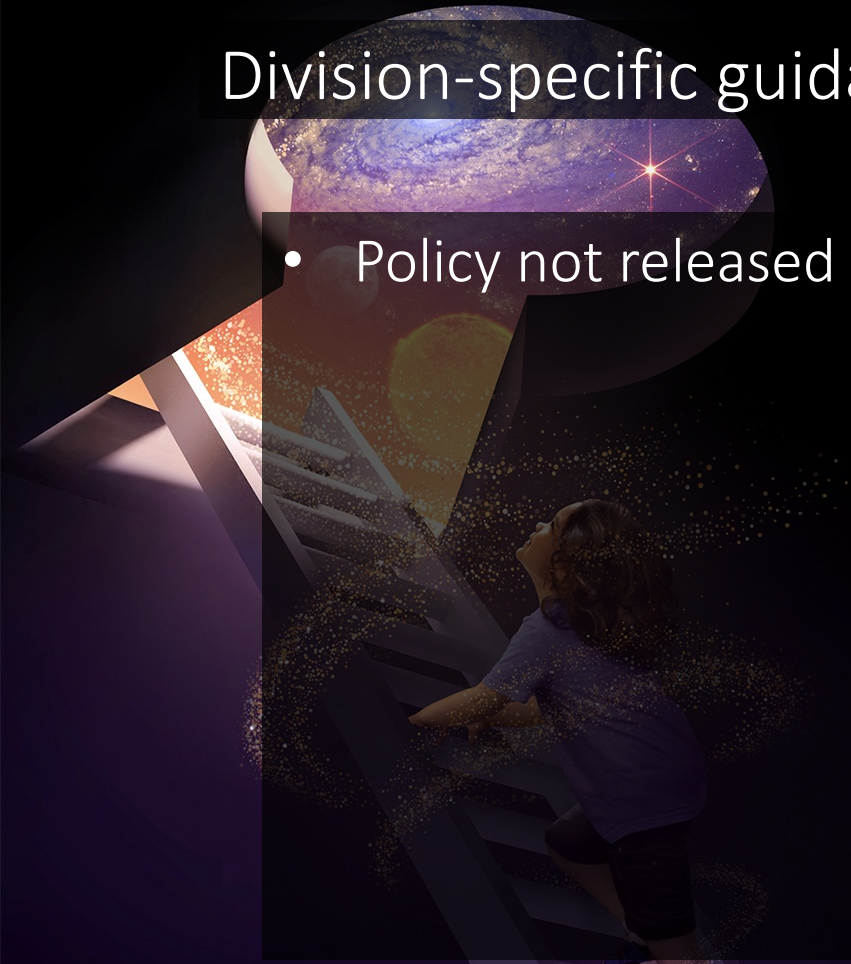
Division-specific guidance — Earth Science

- Software must be shared with a permissive license if possible
- The GNU General Public License (GPL) is specifically forbidden, as it is considered a “restrictive” license, in contrast with the “permissive” license requirement

Division-specific guidance — Heliophysics

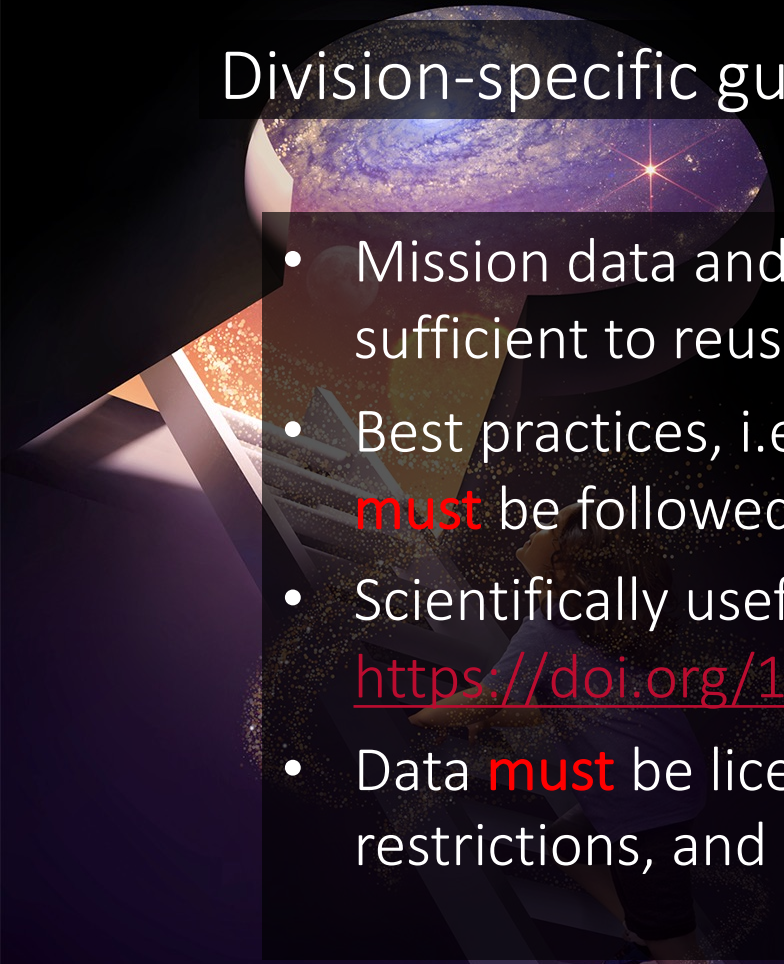
- Mission data **must** be released within 6 months
- Data providers **must** create SPASE records for DOI generation with the SPASE Metadata Working Team: https://hdrl.gsfc.nasa.gov/smwt_home/smwt_index.html
- Scientifically useful data **must** follow FAIR principles: https://doi.org/10.1038/sdata.2016.18
- Data **must** be licensed under CC0 if there are no other restrictions, and software **must** use a permissive license
- All data **must** be shared in an archive, not solely as supplemental information

Division-specific guidance — Planetary Science



- Policy not released yet

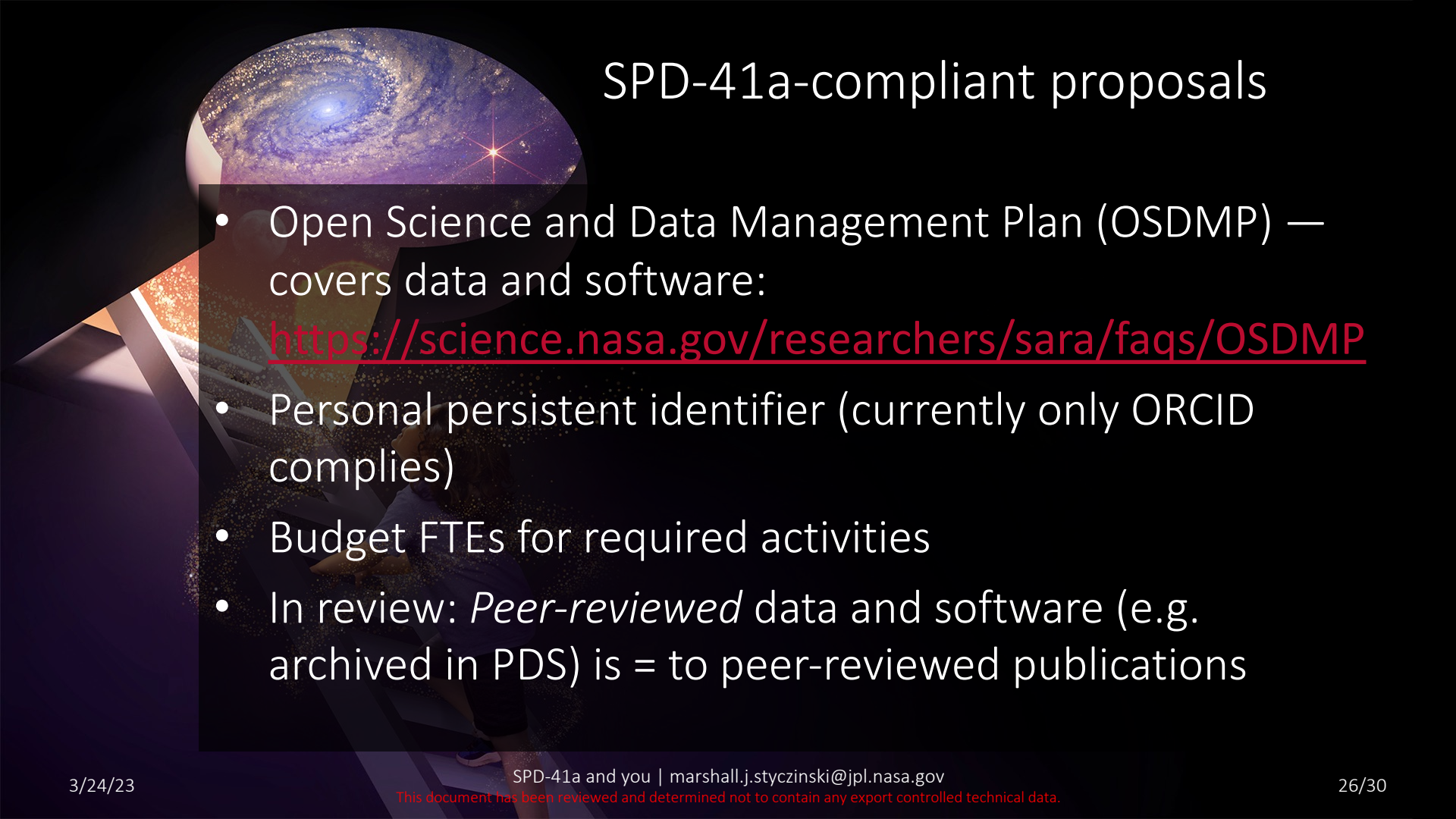
Division-specific guidance — Astrophysics



- Mission data and software **must** include documentation sufficient to reuse and reproduce analysis of raw data
- Best practices, i.e. https://arxiv.org/pdf/2106.01477.pdf, **must** be followed as much as possible
- Scientifically useful data **must** follow FAIR principles: https://doi.org/10.1038/sdata.2016.18
- Data **must** be licensed under CC0 if there are no other restrictions, and software **must** use a permissive license

Division-specific guidance — Biological and Physical Sciences

- Mission data **must** be released within 6 months and within 1 month of validation
- Corrections and recalibrations of archived data **must** be submitted as soon as they are available
- All public presentations (incl. conferences) **must** have slides etc. published via a designated NASA repository
- Project status reports and final reports **must** be reported in Task Book: https://taskbook.nasaprs.com/



SPD-41a-compliant proposals

- Open Science and Data Management Plan (OSDMP) — covers data and software: https://science.nasa.gov/researchers/sara/faqs/OSDMP
- Personal persistent identifier (currently only ORCID complies)
- Budget FTEs for required activities
- In review: *Peer-reviewed* data and software (e.g. archived in PDS) is = to peer-reviewed publications

Personal suggestions for SPD-41a compliance

- Develop software on GitHub, openly if possible
- Archive .zip of software and/or data on Zenodo
- Write single-use software (e.g. making publication figures) in scripts, not command prompts
- Use an OSDMP template!

Doesn’t open science put me at a disadvantage?

- You are the expert in your software! Make yourself available—you know best how to use it!
- New requirements to attach a DOI, license, and cite data and software work in your favor. Tips:
  - When allowed, use a license like GPL-3.0 that requires attribution, over MIT, which does not.
  - Give suggested wording for acknowledging use of your software
- Conducting science openly gives you a long paper trail

SPD-41a and open science

- Half of science is sharing it with the world
- SPD-41a is borne out of a cultural shift:
  - There is incentive to build on each other's software, rather than compete, when it's just easier to cite and use what's already there
  - Creating and maintaining software takes time. Many of us stand to gain a lot from better recognition of our investments there.
  - The Year of Open Science (2023) is partly about incentives for open science! Read more at https://open.science.gov/
- Check out the README on the TOPS GitHub! It contains loads of resources: https://github.com/nasa/Transform-to-Open-Science

Questions?

Jet Propulsion Laboratory
California Institute of Technology

jpl.nasa.gov

marshall.j.styczinski@jpl.nasa.gov
https://github.com/nasa/Transform-to-Open-Science

Thanks to Steve Crawford for suggestions!

This work was supported by the NASA Postdoctoral Program, administered by Oak Ridge Associated Universities under a contract with NASA (80HQTR21CA005).

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA (80NM0018D0004).

TOPS logos credit Tanner Boeger: https://doi.org/10.5281/zenodo.5225075

Bonus: More resources

- Official SPD-41a web pages: https://science.nasa.gov/researchers/science-data/science-information-policy (includes a nice FAQ and links to division-specific policies and guidance)
- Open-source science guidance for SMD-funded researchers (pdf) — more info about how to comply with SPD-41a (also available in a dedicated GitHub: https://github.com/nasa/smd-open-science-guidelines)
- Further questions: send to HQ-SMD-SPD41@list.nasa.gov