Proposal Summary

We propose to create a narrative-driven, interactive Jupyter Book that will serve as a self-paced course in open science data, tools, and techniques. The content will be targeted to learners ranging from interested laypeople to experienced scientists. It will center skills, data, and workflows that are highly relevant to research in planetary science and astronomy, and help users gain multiple competencies necessary to build the open science ecosystem and facilitate scientific research. These include: (1) accessing and analyzing publicly-available NASA data (including cloud-based datasets), (2) using both general and discipline-specific open-source software libraries for data access, manipulation, analysis, and visualization, (3) creating, managing, and sharing open science workflows. The narrative components of the course will be developed by and in partnership with Native people; they will lead learners through engaging hands-on investigations of Native knowledge concepts relating to astronomy and planetary science.

The work will be open, collaborative, and accessible. We will perform all development in a public Github repository, and will develop a set of guidelines and a code of conduct for community contributions. The content will be OS-agnostic, available in both English and Spanish, and will avoid code that requires high-end computer hardware. It will follow relevant portions of the Web Content Accessibility Guidelines to help ensure, as best possible, that it is accessible to users with disabilities. It will also incorporate recommendations from advocacy groups to improve usability for neurodivergent users. The content will be completely free and released under a permissive open-source license.

This project and deliverable are well-aligned with the SMD scientific vision, OSSI, and TOPS project. Specifically, it will teach learners to source observational data from institutions within NASA's Planetary Science and Astrophysics Divisions, key to the SMD goals of "advanc[ing] scientific acknowledge of the origin and history of the solar system" and "discover[ing] the origin, structure, evolution, and destiny of the universe." It will leverage the capabilities of and teach fundamental skills for both open source software and data repositories, increasing open science literacy and encouraging open, transparent, accessible, and reproducible science per the objectives of both OSSI and TOPS. The incorporation of Native narratives, emphasis on accessibility, plan for dual-language support, and open design and development with substantial opportunity for community feedback further address the TOPS objective to broaden participation in SMD-funded research by historically underrepresented communities.

Proposal: "Knowing the Sky: Building Open Science Skills through Native Knowledge Practices"
Program: NASA ROSES22 F.14 TOPST

Synopsis

We propose to create a narrative-driven, interactive Jupyter Book that will serve as a self-paced course in open science data, tools, and techniques. The content will be targeted to learners ranging from interested lay people to experienced scientists and will use Python as its core programming language. It will center skills, data, and workflows that are highly relevant to research in planetary science and astronomy and help users gain several types of competency necessary to build the open science ecosystem and facilitate scientific research:

- accessing and analyzing publicly-available NASA data (including cloud-based datasets)
- using both general and discipline-specific open-source software libraries for data access, manipulation, analysis, and visualization
- creating, managing, and sharing open science workflows.

The narrative components of the course will lead learners through exciting hands-on investigations of Native knowledge concepts relating to astronomy and planetary science. Scientific practice often asserts that there is a single valid way of knowing: that of dominant Western culture. Drawing on Native narratives will work to resist this practice by celebrating and validating other modes of knowledge, helping to Indigenize scientific culture. There is no simple way to undo the effects of centuries of oppression, but we believe that **attempts to attract Native scientists to SMD-funded research that are not paired with efforts to Indigenize and decolonize open science training are misguided and potentially harmful**. It will also help non-Native users build cultural competencies along with open science skills.

The work will be open, collaborative, and accessible. We will perform all development in a public Github repository, and will develop a set of guidelines and a code of conduct for community contributions. The content will be OS-agnostic, available in both English and Spanish, and will avoid code that requires high-end computer hardware. It will follow relevant portions of the Web Content Accessibility Guidelines [1] to help ensure, as best possible, that it is accessible to users with disabilities. It will also incorporate recommendations from advocacy groups to improve usability for neurodivergent users. The content will be completely free and released under a permissive open-source license. **An interactive, highly-accessible course on open planetary science and astronomy techniques that highlights NASA data sources, helps open science to members of marginalized communities, and breaks down technical barriers to scientific discovery is essential to the goals of the TOPS program.**

Goal 1: Building open science skills

Historically, many digitized astronomy and planetary science workflows have relied on closed-source tools and languages like MATLAB, ENVI, etc., largely owing to lack of other options. **There are now viable open-source alternatives.** Due to the high popularity of Python among scientists, its rich ecosystem of open-source libraries (many oriented towards planetary science and astronomy), and its tight integration with a variety of interactive media (particularly Jupyter Notebooks), this effort will use Python as its core language.

<u>Goal 2: Facilitating open science adoption</u>

Open science software and techniques are increasingly popular in science. For instance, a recent anonymous survey found that Python (an open-source language) was the most commonly-used language (at 70% of respondents). [2] **However, the adoption of open science software and techniques among scientists is far from complete.** The authors of the survey noted that Python users were likely overrepresented among respondents, and the runner-up languages (IDL and MATLAB) were both closed-source.

Figure 1. *"""The frequency of the ears and eyes is vary narrow" "The whispering of our ancestors..." is always present. It is our spirit that "binds us with place and land" "You must always come home to were reservoirs of strength are""' Artwork and poetry by Shandin Pete (2019), used with permission*

Better education can promote adoption. Many survey respondents cited a steep learning curve with Python, due in part to the limited availability of educational materials. Respondents also cited simply opening and acquiring relevant planetary and astronomical data as major pain points. **There is a scarcity of educational resources for scientists who wish to gain expertise in open science methods.** The recent Planetary Data Ecosystem Independent Review Board report noted that there is a gap for "intermediate to advanced" level training materials for domain-specific programming. [3]

Promoting further adoption of open-source languages and open science techniques is a linchpin goal of the TOPS program. Unfortunately, the educational ecosystem for open science is scattered and underdeveloped. This effort will develop consolidated training materials to guide learners towards proficiency in applying open source to astronomy and planetary science.

<u>Goal 3: Honoring Native culture</u>

Both the project PI and our education consultant are enrolled members of recognized tribes, and are deeply concerned with accurately representing cultural practice and avoiding the presentation of a monolithic "Native culture" that hides diversity and erases differences between tribes. They are also highly aware of the importance of ensuring that when traditional knowledge is shared, it is done so with respect and permission. **We will draw on Native knowledge they are expert in, and preferably, knowledge that is already public:** their own heritage practices (including personal communications with members of their tribes) and public data from archives of anthropological studies of closely related tribes in the early 1900s [e.g. 4-5].

When describing the Native knowledge we propose to use in this effort, we use phrases like "Native oral tradition" and "Native knowledge practice." This terminology is intended to express two core beliefs: (1) **We assert that *both* Native cosmologies and Western scientific disciplines are legitimate bodies of knowledge.** This is the basis of the Mi'kmaq term "two-eyed seeing"; *combining* Indigenous and Western scientific practices can produce novel, holistic perspectives and solutions. However, most contacts between Native traditions and Western-style scientific disciplines have not occurred on equal footing. Native tradition has historically been subject to a coloniality of knowledge with results ranging from misappropriation to erasure and destruction [6]. Western science, in its modern form, grew up alongside colonialism. Scientific knowledge has often been treated as a commodity, extracted, like other commodities, at the expense of Native people; and has in turn been used to facilitate and legitimate colonial projects [7]. Terms like "myth" and "folklore", while not pejorative, risk reenacting these colonial encounters, relegating Native tradition to the role of, at best, respectfully-framed decoration. (2) **Many Native knowledge practices and bodies of oral tradition are living and active.** Using the lens of "myth" risks reinforcing the persistent, harmful narrative that Native people are solely beings of the past, rhetorically erasing Native people and cultures that exist today.

<u>Goal 4: Building a diverse scientific workforce</u>

Native Americans are underrepresented in science, making up over 2% of the US population [6], but only 0.3% of the STEM workforce [9]. Numerical underrepresentation and cultural exclusion form a vicious cycle. **Incorporating Native knowledge practices into scientific education works towards breaking this cycle by demonstrating that Native people and culture belong in science.** Programs supportive of Native culture and knowledge practices improve retention of Natives in STEM education [10]. **This supports the TOPS program's goal to further open SMD-funded research to members of historically underrepresented communities.**

<u>Deliverables</u>

The primary deliverable is a Jupyter Book that works as a guided self-study course for open science skills. It will consist of interactive Jupyter Notebooks and contextualizing Markdown documents. **Each section of the Book will pair essential skill clusters with a set of Native narratives and/or cultural practices, guiding learners through techniques to use open data with open tools to explore questions that are at once scientific, aesthetic, and cultural.** While individual sections of the Book may be useful on their own, it is not intended as a reference manual. Learning objectives will be structured like a class. **Each section will be cumulative, allowing learners to build on skills they gained in earlier sections. Interactive coding exercises will serve as self-assessments to help facilitate this process of skill consolidation.**

This project and deliverable are well-aligned with the SMD scientific vision, OSSI, and TOPS project. The Book will teach fundamental technical skills relevant to planetary scientists and astronomers across a variety of subdisciplines. Specifically, it will teach learners to source observational data from institutions within NASA's Planetary Science and Astrophysics Divisions, key to the SMD goals of "advanc[ing] scientific acknowledge of the origin and history of the solar system" and "discover[ing] the origin, structure, evolution, and destiny of the universe." It will not assume familiarity with specific tools, data sources, or libraries. Neither will it assume a specific career stage or professional goal. It will aim to be useful to learners ranging from upper-division undergraduates to early-career professionals to interested laypeople, and leverage and teach the fundamentals of both open source software and data repositories, increasing open science literacy and encouraging open, transparent, accessible, and reproducible science per the objectives of both

OSSI and TOPS. The incorporation of Native narratives, emphasis on accessibility, plan for dual-language support, and open design and development with substantial opportunity for community feedback further address the TOPS objective to broaden participation in SMD-funded research by historically underrepresented communities.

We are confident in our curricular vision, but note that the specifics described below are subject to change for several reasons: (1) Full curricular development is a substantial effort and a major portion of the work plan. (2) We plan to solicit community input. Two Notebooks are earmarked for community contributions, but we wish to be responsive to contributions on all course content. (3) We may modify content to add newly-available libraries or data sources. (4) The PI and education consultant may determine, after further consultation with cultural experts in their respective tribes, that some narratives should be presented differently.

Notebook	Skills	Libraries Introduced
1	Getting data from NASA cloud sources. Managing data locally.	*astroquery, boto3, os, pathlib, shutil, requests*
2	Opening tables and arrays. Indexing and array visualization.	*numpy, pandas, pdr, matplotlib.pyplot*
3	Creating plots. Working with time and coordinates.	*astropy, matplotlib.figure*
4	Parsing text with regular expressions. Producing descriptive statistics. Model fitting.	*re, scipy*
5	Working with SPICE. Building symbolic expressions.	*sympy, spiceypy*
6	Image processing: filtering and masking images. Similarity analysis using computer vision.	*cv2, scikit-image*
7	Preprocessing data for AI/ML pipelines. Using classification and decomposition models.	*scikit-learn*
8	Creating and versioning GitHub repositories. Managing data on AWS S3.	*boto3* (advanced usage), *pillow*
9-10	TBD (based on community input)	TBD

Table 1. Summary of course content by notebook.

<u>Course outline</u>

<u>Preliminary</u>

GitHub does not provide code execution capabilities, and environment setup skills are core to developing self-efficacy with open science tools, so the course will start with a section on using *conda* to create an environment for executing the Book.

<u>Section 1: Gathering, Ordering, and Drawing the Sky</u>

Ethnographies and oral traditions demonstrate that Salish peoples had advanced astronomical knowledge long before colonization. One significant asterism in this body of knowledge is a group of three stars that never change their "position of rising." Ongoing research has determined that Salish traditions often link these three stars to a narrative of a "canoe with three brothers," and that the stars that represent the brothers are the stars in the "belt" of the Western constellation Orion: from the viewpoint of Salish territory, they always rise halfway between the place that moon rises and the place that the sun rises. This section guides users through the puzzle of identifying these stars based on ethnographic records of the narrative of these three brothers, allowing them to explore two ways of seeing this asterism (Orion/Canoe).

Notebook 1: Gathering the Sky – NASA and its affiliates, such as the Space Telescope Science Institute (STScI) and the nodes of the Planetary Data System (PDS), hold one of the richest collections of open data in the world, increasingly available through cloud platforms. However, limited knowledge of the techniques to most effectively access these data presents a barrier to use. This Notebook will teach essential data access skills by guiding learners through acquisition and organization of data relevant to investigating possible celestial referents of the anthropological account. Using the HTTP library *requests*, the AWS API wrapper library *boto3*, and the domain-specific cross-platform API wrapper library *astroquery*, they will learn how to fetch products from a variety of NASA sources while gaining knowledge about the cloud, data archives, and fundamental Internet technologies like HTTP. Next, they will learn to organize the data locally using core modules of the Python Standard Library like *pathlib*, *os*, and *shutil*.

Learning objectives: Familiarity with key cloud and Internet technologies, the distinctions between them, and their uses; knowledge of basic archive structures; critical awareness of how archives differ from bodies of oral tradition and how scientists might incorporate strengths of oral tradition into open science practices; basic syntax of and entry points to the libraries mentioned above; best practices for managing bandwidth, storage, and local data.

Notebook 2: Ordering the Sky – Due to the enormous diversity of scientific data formats, simply "opening up" data—loading it from files into memory and displaying it in meaningful ways—often presents a barrier to many. Fortunately, powerful open-source tools exist to mitigate the problem. This Notebook will present workflows for data reading, display, and manipulation to investigate the locations of the sun, moon, and stars, with applications to original research.

Learning objectives: awareness of Python libraries for representing basic scientific data formats, e.g. *numpy* for arrays and *pandas* for tables; knowledge of basic methods of use of these libraries; ability to use the Planetary Data Reader (*pdr*) library to read data; capacity to perform indexing and slicing operations on *pandas DataFrames* and *numpy ndarrays*; knowledge of the *matplotlib* library, and familiarity with techniques for visualizing arrays with it (e.g., the *imshow* function).

Notebook 3: Drawing the Sky – Visual representations of observational data are crucial conceptual tools in science, and Python offers many options for visualization tools. This Notebook will guide learners towards competence with visualization techniques by exploring how the anthropological account might express the path of celestial bodies across the sky, paired with an introduction to techniques for manipulating coordinate systems, keeping time, and calculating geometry.

Learning objectives: Ability to create plots in *matplotlib*; familiarity with core *astropy* library modules, including *coordinates* and *time;* awareness of conceptual similarities between data representation techniques and and expressive elements of oral knowledge practices.

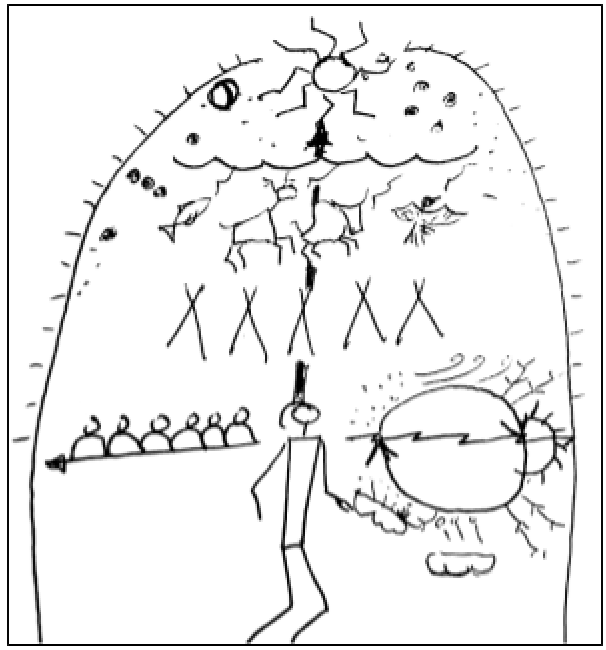

Figure 2. Artwork by Shandin Pete (2021), used with permission

Section 2: Naming and Measuring the Sky

While the Salish still possess rich astronomical knowledge, much of it has been lost to the physical and cultural violence of colonialism. Even at the time these narratives were recorded, their Salish sources described some of them as distant memories. One of these lost narratives concerned Elk and Antelope. As they were racing, they turned into stars and continued their race in the sky. A similar narrative is referenced in other accounts. In this section we guide users to explore the hypothesis that the Elk and Antelope reference visible planets: irregularly moving celestial objects that cross paths, back and forth.

Notebook 4: Naming the Sky – The challenges scholars of Native traditions experience when researching knowledge of this type are closely related to challenges in scientific research like finding references in documentation or synthesizing metadata. Open-source offers powerful text-processing and statistical tools to facilitate these tasks. This Notebook guides learners towards mastery of these tools by pairing searches of interview texts with metadata organization, looking for statistical correspondences between narrative elements and astronomical data. These techniques are generalizable to many other analysis and modeling problems in scientific research.

Learning objectives: Familiarity with basic text parsing techniques; ability to compose regular expressions ("regex"); competence with descriptive statistical functions from the *scipy.stats* module; ability to use functions from the *scipy.optimize* module to perform linear model fits.

Notebook 5: Measuring the Sky – In Notebook 4, learners cross-referenced multiple historical accounts to find more information on the "racers" narrative and determined some plausible celestial referents. More detailed planetary dynamics and ephemeris data could help explore references. There are a number of "thick wrappers" that help users merge non-Python tools and libraries into Python workflows. This Notebook will showcase this, along with an important portion of NASA's data holdings, by introducing users to the *spiceypy* library, a wrapper for NAIF's SPICE toolkit for accessing and using ephemeris data. It will also introduce *sympy*, a symbolic mathematics library that synergizes with the numerical analysis libraries introduced in previous Notebooks. Users will build orbital dynamics expressions and compare them to historical ephemeris to find relationships between celestial bodies and the racers in the Salish narrative.

Learning objectives: Understanding of the "wrapper" concept; ability to find and download SPICE kernels; ability to use *spiceypy* to load ephemeris data from SPICE kernels; basic facility with building *sympy* expressions and converting them to numerical functions.

Section 3: Seeing, Describing, and Sharing the Moon

The PI's tribal traditions are tightly connected to the tides and seas; they know the Moon as a Grandmother. Female members of the tribe have a special connection to Grandmother Moon, and seek guidance and wisdom from her during Full Moon Ceremonies. (Moon phases are also important in many other Native practices across many tribes, although different tribes hold many different sorts of knowledge about the Moon.) In our first Notebook in this section, we guide users through a variety of exercises that explore what it means for the Moon to be "full."

Salish oral traditions describe the selection of the being that would become the Moon. In this narrative, a bullfrog clings to his face as he becomes the Moon, eventually becoming the dark spot we now see on the surface. In our second Notebook, we explore ways of visualizing this frog.

Notebook 6: Seeing the Moon – Western science offers a strict definition of "fullness"—percent illumination as determined by idealized orbital geometry—and the tools introduced in Sections 1 and 2 of the book are adequate to determine "fullness" in this sense. However, the phase of the Moon is not ontologically reducible to geometric states. Ways of knowing that incorporate visual experience and site-specificity offer other paths to understand the Moon. Roughly analogous practices are central (if somewhat abject) parts of mainstream Western science: most scientists investigate their objects of study in large part via expressive visual representation. In recent years, *automated* seeing, or "computer vision", has become increasingly important in science. This Notebook will guide learners through an investigation of multiple meanings of "fullness" by introducing them to features of two important computer vision libraries (*scikit-image* and the *cv2* wrapper to OpenCV). They will perform morphological analysis on a range of lunar images, compare these to ephemerides, and determine when, where, and *how* the Moon might be *full*.

Learning objectives: critical awareness of multiple ways of knowing "phase"; general familiarity with the capabilities of *scikit-image* and *cv2*; ability to create masks and apply morphological filters with *scikit-image*; ability to perform basic similarity analysis with *cv2*.

Notebook 7: Describing the Moon – This Notebook will teach users to draw on the machine learning capabilities of the scientific Python ecosystem to develop ways of seeing the frog (referenced in Salish oral traditions) that complement the morphological analysis techniques discussed in Notebook 6. Learners will build a supervised classifier to investigate what precise *species* of bullfrog might best match features of the lunar surface and then build an unsupervised model to explore how this restricted sense of "best" might be expanded: i.e. in what other ways might a statistical model learn to "see" a frog on the Moon? These techniques are extensible to many domain-specific questions in planetary science and astronomy.

Learning objectives: Familiarity with basic capabilities of the *scikit-learn* library; ability to build a preprocessing pipeline for machine learning; awareness of major types of machine learning strategies; ability to build, train, and score simple classification and decomposition models.

Notebook 8: Sharing the Moon – By this point in the course, learners will have gained familiarity with many open science tools and data sources. However, the ability to find and use open resources is only part of the practice of open science. The ability to share your own resources with the community is equally important. This Notebook will build on prior skills by familiarizing learners with core resources for distributing software and data. They will learn new *git* commands and GitHub functionality to create a repository and share the code they wrote in the previous two Notebooks. They will create a free-tier AWS account, make an S3 bucket, and use *boto3* to place model-generated images in that bucket to be viewed by others.

Learning objectives: ability to create and version GitHub repositories; knowledge of SSH keys and other authentication methods; awareness of capabilities and limitations of object storage systems; ability to create S3 buckets and store objects in them with *boto3*.

Section 4: Placeholder for emerging needs

We have scoped for two additional notebooks with topics and content TBD based on community suggestions and contributions, or relevant technological advances during the period of performance. We predict that community contributions may focus on more specialized or advanced skills, so it will be most appropriate to place them in a final section. If this is not the case, we might fold them into one of the previous sections instead.

Work Plan

We have developed a plan of work and project estimate / budget using a method of task decomposition derived from industry best practices and adapted for academic research. **This process yields estimates that are both reasonable and sufficient for the work as scoped.** In brief, the project is broken into discrete tasks, with an estimate generated separately for each and then scaled by an uncertainty factor based on the team's (substantial) prior experience performing similar work—where less experience means higher uncertainty / scale. The method is described in complete detail in [11]. **We have determined that 0.5 FTE of effort is required to ensure project success**, which has been split across investigators and project years. This effort does not include the level of effort from the contracted teaching consultant or translator. The tasks are described in detail below and also represented in a Gantt chart (fig. 3). The investigators will share effort in nearly all tasks, at the direction of the PI, who will also perform project management duties and ensure that the work proceeds towards success and that funds are properly managed. The investigators will work remotely at their home offices. The team will remain in frequent contact via email, chat, and videoconferencing. The investigators have collaborated remotely, with success, on multiple projects over several years. All investigators have access to work space, computers, and high-speed internet; no other facilities or equipment are required.

Task 1: Background Research and Framing

This task encompasses high-level course design, including technical and non-technical content. We will perform most of this task in the first six months of the project. Some effort will continue through the second year, based on feedback from testers, the teaching consultant, tribal elders and members, or the community at large.

This project team is committed to respecting and ensuring willing participation of Native peoples, so this design effort includes a significant research component to ensure that we draw on Native cultural traditions in our materials in an appropriate and accurate way. The PI and teaching consultant will lead efforts to communicate with elders and regulatory bodies from their respective tribes to secure explicit permissions and acquire additional contextualizing knowledge. We have had preliminary discussions with other members of these tribes prior to submitting this proposal, but if we encounter any concerns related to use or description of specific content, we have identified alternative narratives/traditions we might use in its place. This task also includes a textual research component to deepen our understanding of knowledge sourced from public texts.

This task includes graphical and curricular design elements. We will produce a detailed outline of learning objectives per Notebook and interactive self-assessments for these objectives, and determine how information should be split between them and contextualizing Markdown documents. We will develop a graphical style and shared Notebook template for the Book and evaluate our information and graphic design choices for accessibility, including compliance with the Web Content Accessibility Guidelines. Although we do not believe an equivalent course currently exists, we will attempt to synergize with existing resources to the extent practical. In particular, we will maintain an awareness of the OpenCore curriculum when it becomes available.

<u>Task 2: Material Composition</u>

This task encompasses implementing the course design developed in Task 1. Effort on this task will peak in Q3 of project YR1 after the bulk of curricular design and background research is complete, and will ramp down in Q3 of YR2 as we move into the testing and refinement phase. Curating data products for use in training materials will be the most time-consuming part of this task. We plan to draw largely from pre-existing cloud-accessible NASA data sets (e.g. MAST and PDS archives), but will need to carefully select products that facilitate good demonstrations of tools and techniques. However, in a few cases—notably the frog image training set in Notebook 7— we will need to use non-NASA data and compile bespoke data sets.

This task includes composing draft content (including text and code) for the Jupyter Book, along with development of supporting material to help potential users discover and use the course material. Supporting material will reference existing documentation when possible. This also includes several communication subtasks, which will be facilitated by drafting course materials in a public GitHub repository. We will solicit and incorporate community input and feedback as described in the Open Source Science Development Plan. The teaching consultant will review materials as they are being produced to help ensure they appropriately and accurately honor Native knowledge practices. We will hire a Translator to produce a Spanish-language version of course materials, and will iterate with them on translation choices, especially regarding translations of computer code and Indigenous vocabulary.

<u>Task 3: Coordination Meeting</u>

The PI will attend the annual 4-day meeting in Washington, DC each year of the project to coordinate effort with other TOPST teams. At the time of proposal preparation, dates for this meeting had not been determined, so we have assigned placeholder dates to this task.

Task 4: Testing and Delivery

This task encompasses content and layout refinements, quality assurance, copyediting, and final delivery of materials in the final six months of the project. Our Data Analyst (DA) will undertake a testing effort, reviewing both technical features (i.e. whether things run properly) and content. We will also solicit testers from the wider community through a variety of informal and formal communication networks within the disciplines of planetary science and astronomy, including but not limited to Planetary Exploration Newsletter (PEN), OpenPlanetary, social media platforms, the Spacecraft Planetary Imaging Facility (SPIF), various Space Grant Consortia (SGC), and the TOPS network. This solicitation is distinct from our solicitation for community design input (in Task 2): because these are educational materials, the best *testers* and the best *designers* are probably not the same people. We will use feedback from our DA and volunteer testers to help identify opportunities for improvement (e.g. inadequate descriptions, additional assessments/exercises), and conduct an iterative design step to implement these improvements.

We will also deploy the Jupyter Book to a cloud execution service (e.g. Binder) to make the training materials accessible to learners who cannot or prefer not to execute the Book locally. Finally, we will package and deliver the Book and supporting materials (including technical documentation) to a TOPS-owned GitHub repository. Technical details of delivery are discussed more completely in the attached Open Source Science Development Plan.

			YR1 (6/1/23 - 5/31/24)						YR2 (6/1/24 - 5/31/25)					
Task Number	**Task Description**	**FTE Est.**	**June - July**	**Aug - Sept**	**Oct - Nov**	**Dec - Jan**	**Feb - Mar**	**Apr - May**	**June - July**	**Aug - Sept**	**Oct - Nov**	**Dec - Jan**	**Feb - Mar**	**Apr - May**
1	Background & Framing	0.15	0.02	0.02	0.03	0.03	0.01	0.01	0.01	0.01	0.01			
1.1	Framework & Template Development	0.03	0.01	0.01	0.01									
1.2	Indigenous Narrative Background	0.09	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01			
1.3	Curricular Design	0.02			0.01	0.01								
1.4	Accessibility Design	0.01				0.01								
2	Material Composition	0.2	0.002	0.003	0.003	0.024	0.034	0.034	0.029	0.029	0.024	0.015	0.003	
2.1	Data Curation	0.07				0.02	0.02	0.015	0.01	0.005				
2.2	Supporting Material Composition	0.01										0.01		
2.3	Notebook Drafting	0.08					0.01	0.015	0.015	0.02	0.02			
2.5	Contractor Communications	0.02		0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	
2.6	Facilitating Community Input	0.02	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002		
3	Coordination Meeting*	0.05			0.025						0.025			
4	Testing & Delivery	0.09									0.005	0.015	0.035	0.035
4.1	User Testing	0.02									0.005	0.005	0.01	
4.2	Testing & Refinement	0.05										0.01	0.025	0.015
4.3	Packaging & Delivery	0.02												0.02
Total		**0.49**	0.022	0.023	0.058	0.054	0.044	0.044	0.039	0.039	0.064	0.03	0.038	0.035
	* Dates of travel are notional.													

Figure 3: Gantt chart showing the distribution of work effort across time and tasks.

Citations

[1] Web Content Accessibility Guidelines (WCAG) 2.1, 2018. URL https://www.w3.org/TR/WCAG21/.

[2] St. Clair, M., Million, C., Aye, M., Burrell, A., Hare, T., Laura, J., Annex, A., Barrineau, M., 2019. Planetary Science Software Survey. URL https://github.com/MillionConcepts/software-survey-2019

[3] Final Report of the Planetary Data Ecosystem Independent Review Board, 2021.

[4] Boas, F., Teit, J.A., Farrand, L., Gould, M.K., Spinden, H.J., 1917. Folk-tales of Salishan and Sahaptin tribes. American folk-lore society, Lancaster, Pa., New York.

[5] Teit, J.A., 1912. Mythology of the Thompson Indians. Memoirs of the AMNH ; v. 12, pt. 2.

[6] Quijano, A., 2007. Coloniality and Modernity/Rationality. Cultural Studies 21, 168–178. https://doi.org/10.1080/09502380601164353

[7] Smith, L.T., 1999. Decolonizing methodologies: research and indigenous peoples. Zed Books ; University of Otago Press ; Distributed in the USA exclusively by St. Martin's Press, London ; New York : Dunedin, N.Z. : New York.

[8] US Census Bureau, 2020. 2020 Census Results. Census.gov. URL https://www.census.gov/2020results.

[9] Sandia National Laboratories, 2016. The Current State of STEM Education for American Indian and Alaska Native Communities.

[10] Estrada, M., Young, G., Flores, L., Hernandez, P.R., Hosoda, K.K., DeerInWater, K., 2022. Culture and Quality Matter in Building Effective Mentorship Relationships with Native STEM Scholars. BioScience 72, 999–1006. https://doi.org/10.1093/biosci/biac064

[11] Million, C., 2022. a practical guide to research and software project estimation. URL https://github.com/MillionConcepts/software_project_management

Open Source Science Development Plan

This project team is committed to promoting, performing, and facilitating fully open and reproducible research. All materials will be developed openly in a public GitHub repository and will carry a BSD-3 and/or a permissive Creative Commons license, as appropriate. Anyone who is interested in contributing to this project in-progress may do so, subject to the contributor guidelines and code of conduct that will appear in the working repository from the start of the project. The exact text of these guidelines is TBD but will draw on documents and best practices from major open-source scientific computing projects, such as astropy and numpy. We will specifically solicit community feedback regarding tools and techniques for section 4 (as described in the course outline above and Task 2) as well as testing and user feedback to supplement our budgeted testing by our team-internal Data Analyst (Task 4). These solicitations will be made through a variety of informal and formal communication networks within the disciplines of planetary science and astronomy including by not limited to: Planetary Exploration Newsletter (PEN), OpenPlanetary Slack, social media platforms (LinkedIn, Twitter), Spacecraft Planetary Imaging Facility (SPIF), various Space Grant Consortia (SGC), and the TOPS network. We will accept community feedback using Github issues or through the communication networks they were solicited through and incorporate community input using Github pull requests.

Software Management

The primary work product will be an interactive Jupyter Book consisting of a series of Jupyter notebooks interleaved with explanatory markdown documents. This content will be publicly hosted in one or more public Github repositories affiliated with the proposing institution, from the first day of the project, with a BSD-3 license. In the spirit of the program, we will also constrain any developed software to use open source software and open data dependencies. Development will occur continuously and in the open on that platform. We plan that the final software will be accompanied by information on how to create virtual environments in which the software can be successfully executed, including Python environment definitions or Docker build files. Because there does not yet exist any robust solution for archiving executable software, the inclusion of thorough documentation on build environments is intended to future-proof the usability of the interactive Jupyter Book. This project is not intended to produce original scientific results, so the goal is not reproducibility or "validation of published results," but continued usability of the software for training purposes for many years into the future. At the end of the period of performance, the Jupyter Book and related materials, data, and documentation can be transferred to the TOPS Github through a number of straightforward mechanisms, and we will work with the TOPS team to make sure this transfer goes smoothly.

Data Management

The project's primary deliverables do not include "data" in the usual sense the term is used in scientific research. Most data referenced in the Book will be dynamically retrieved by users from preexisting open and internet- or cloud-retrievable data sets. However, we are likely to produce derived/reduced/preprocessed versions of some existing data products—for example, a "cleaned" set of frog image data that is easy to manipulate with scikit-learn pipelines—in order to facilitate more direct pursuit of particular learning objectives. We will make any such products freely available to the public during the period of performance on a platform such as Google Cloud Storage or AWS S3 with a permissive Creative Commons license. We do not yet have a good estimate for the volume of such data, but we expect it to be less than a gigabyte in total. During project closeout, we will make arrangements with TOPS to either transfer these products to the TOPS Github or other NASA-managed storage or replace them with suitable alternatives already in open data sets. At minimum, data can be archived with Zenodo.