NASA Science Mission Directorate
Research Opportunities in Space and Earth Sciences – 2022
NNH23ZDA001N-TOPST
F.14 TRANSFORM TO OPEN SCIENCE TRAINING
Abstracts of Selected Proposals

The National Aeronautics and Space Administration (NASA) Science Mission Directorate solicited proposals supports proposals to advance open science literacy for all who conduct research relevant to NASA's SMD through training and workshops targeting audiences from undergraduate students to established scientists and managers. The solicitation requested investigations that would conduct summer schools, support virtual cohorts, and expand the content of material to advance Open Science in the Science Mission Directorate. NASA received a total of 34 proposals and has selected 16 for funding at this time. The total funding to be provided for these investigations is approximately $6.5 million over thirty-six months. The investigations selected are listed below. The Principal Investigator, institution, investigation title, and project summary are provided. Co-investigators are not listed here.

Laura Acion/Code For Science And Society, Inc.
Ciencia Abierta Accesible: Community-Based Teaching of the TOPS OpenCore Online in Spanish
22-TOPST22-0016

Objectives
- Train at least 5 cohorts per year for 3 years. Each cohort will receive 6 weeks of online training covering the 5 NASA TOPS OpenCore modules in Spanish.
- Steward a supportive and inclusive Spanish-speaking online community around NASA TOPS training.

Methods
Rather than short-form training devoid of a strong community around it, the proposed 6-week Virtual Cohorts provide opportunities for learning through hands-on implementation of open science. Community-based training allows for continued engagement beyond the 6 weeks of the program.

Open-science-curious researchers, from undergraduates to senior practitioners, are asked to commit 4 hours a week for 6 weeks. Learning materials and synchronous Zoom calls will be in Spanish. Each cohort will include a mix of online training sessions from domain experts to cover all TOPS OpenCore modules, discussion groups to explore nuance, hands-on work applying open principles and techniques to trainee's work, and 1:1 mentoring.

When needed, domain experts will be offered professional-quality live-interpretation from English to Spanish, and/or coaching on pedagogy for online training. Mentors are experienced open science practitioners, with professional coaching training.

Asynchronous activities and community immersion will happen via a Spanish-speaking Slack where cohorts, mentors, experts, organizers, and other open science practitioners will be able to interact before, during, and after the 6-week program.
Offering the TOPS OpenCore in Spanish is essential for lowering the language barrier, therefore allowing the participation and community engagement of Spanish-speaking scientists.

Significance of the proposed work
TOPS Goal: Increasing understanding and adoption of open science." We will advance open science literacy among a wide scientific community. Our proposal has the overarching goal of allowing open and equitable communities to disseminate their work, as it is designed to fully engage and embrace a multiplicity of talents, ideas, and perspectives. Inclusion by design is one of the core actionable principles we advocate and strive to follow. To us, Open science literacy means deep understanding and adoption of approaches and techniques, but also of principles.

TOPS Goal: Doubling the participation of marginalized communities in open science." Spanish-speakers are numerous in the US, and, with virtual cohorts, it is also possible to reach global communities with that demographic and an interest in open science.

The principle of mutual help is centerstage for us. We are experienced at successfully providing scientists with the help, encouragement, and equal access that participants would need to complete OpenCore modules, earning their badge and, therefore, for the Virtual Cohorts in Spanish to be a success.

This application complements another proposal, but is Spanish-language first.

Richard Barry/NASA Goddard Space Flight Center
ETHOS: ExoplaneTs in the epocH of Open Science
22-TOPST22-0018

Problem statement: To understand the array of exoplanetary systems we observe today and, specifically, to understand how habitable planets are formed by such systems, requires information about the mass function of all bound and free floating planets. Fortunately, the complimentary techniques of exoplanet transits and gravitational microlensing, when taken together, are sensitive to the complete mass and separation range of all possible exoplanets - from massive, boiling hot Jupiters skimming the very atmosphere of their host star [1] and planets with atmospheres of super-hot evaporated rock [2] all the way out to Neptune and Pluto analogs and rogue planets, floating through space unbound to any star [3]. These techniques, which are sensitive to the detection of planets in formation environments from the local Milky Way spiral arm all the way to Sagittarius A*, form the observational basis for NASA's flagship Roman Space Telescope (RST) that will complete the census of all possible planets in our host galaxy beginning in the latter half of this decade. RST, with an array of 18 4k x 4k detectors

imaging the sky every fifteen minutes over it's six year mission will produce approximately 1.9 PB of data. How will science meet the enormous challenge of simply locating transiting and microlensing planets in this vast quantity of data? Can we engage a broader community of new researchers in this exciting science?

The Experiment: We here propose a ScienceCore learning module that may be used to deliver instruction in the use of NASA's archival public data together with our existing open artificial intelligence framework to classify lightcurves in the search for new planets. The module will leverage cloud-based tools and compute power to assure that it is accessible to the broadest possible demographic range of new learners including the hearing and visually impaired. ETHOS will provide an introduction to two of the most important modern machine learning techniques: random forests and neural networks, that make identification of these transits and microlensing events well within the understanding and computational capabilities of many more researchers than currently represented.

Relevance: A core activity of science is the exchange of information. The exoplanet community, however, is under-performing its potential because the ecosystem is largely closed. Historically, R1 universities and the English language have been heavily favored in science resulting in gravely rarefied participation. ETHOS directly confronts this dynamic by placing the tools and data needed to conduct research into the hands of the broadest possible demographic. The ETHOS ScienceCore module will be translated into the two most-spoken languages in the United States after English: Spanish and Mandarin. In addition, we will translate the module into Portuguese and assure that it is accessible to blind and deaf persons to permit penetration into the very largest possible demographic.

Urgency: A timely and successful execution of the ETHOS program will open new discovery spaces for astrophysics in the near term and will be a fundamentally enabling component of future open science survey missions such as NASA's flagship Nancy Grace Roman Space Telescope mission due to launch in 2027

Carl Boettiger/University of California, Berkeley
Examining Environmental Justice through Open Source, Cloud Native Tools
22-TOPST22-0002

In the same year as the first manned mission to the moon, racially segregated housing became illegal with the Fair Housing Act. The law would now ban practices known as redlining in which the federal government's Home Owners Loan Corporation dividing cities into areas graded 'minimal risk' to 'hazardous' for home loans based largely on racial and ethnic make-up. But the consequences of such practices are not easily reversed. More than 50 years later, scientists can still see the pattern of the inequalities etched onto those maps even from space. This educational module will seek to introduce students to the open source platforms and tools used to manipulate and analyze NASA's open earth observation imagery through the lens of examining the environmental legacy of redlining

practices in major urban areas. We will introduce key concepts of cloud-native geospatial workflows including STAC, COG, and GDAL Virtual Filesystem to allow students to leverage analysis against terrabytes of NASA data. The module design is also deeply rooted in pedagogical practices established by education research literature to increase engagement and reach an inclusive audience. The module will be provided in both R and Python and be made available in both English and Spanish translations.

Sierra Brown/Million Concepts LLC
Knowing the Sky: Building Open Science Skills through Native Knowledge Practices
22-TOPST22-0010

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.

Tansu Daylan/Washington University
ExoCore: An Open Science Curriculum For Enhanced Reproducibility And Equity In Exoplanet Research
22-TOPST22-0006

Our ability to discover and accurately characterize planets beyond our Solar System is developing fast. Furthermore, exoplanet research is highly relevant to supporting the goals and vision of NASA's Science Mission Directorate (SMD). Specifically, it is aligned with Strategy 1.3 in Science 2020-2024: A Vision for Scientific Excellence and NASA's 2022 Strategic Plan through its Exoplanet Exploration Program (ExEP). This compelling nature of exoplanet research has made it a major driver of contemporary astrophysics. Nevertheless, researchers from unconventional backgrounds and historically under-represented communities can still suffer from barriers to entering this field. Fortunately, the availability of inclusive, up-to-date, and entry-level research materials can make exoplanet research more equitable.

We propose to develop ExoCore: an accessible and comprehensive open science curriculum that covers research on exoplanets and planetary science. ExoCore will be a crucial contribution to ScienceCore that consists of six modules that cover the exploration, visualization, and modeling of large public data sets from NASA-funded missions such as Kepler and the Transiting Exoplanet Survey Satellite (TESS). In particular, it will contain complete workflows in the form of Jupyter notebooks with substantial guiding narration and interactive visualizations that enable active learning. This will make ExoCore easily executable by capacity-building Transforming to OPen Science (TOPS) activities such as summer schools and virtual cohorts. We will actively maintain ExoCore and offer help to all TOPS teams engaging with the material.

ExoCore will introduce the audience to major data sets used in exoplanet research such as the NASA Exoplanet Archive, TESS Objects of Interest catalog, and TESS Input Catalog (TIC). It will focus on the criticality of transparency and reproducibility. We will exemplify the access to and analysis of TESS data on the cloud and maintain a list of open-access tools used in exoplanet research. ExoCore will illustrate how these data and tools can be used to improve our understanding of planet formation and evolution as well as the origin of our Earth and the Solar System. Complementing OpenCore, ExoCore will be available on the TOPS GitHub and reinforce engagement with Open Science. Through the use of tailored color maps, we will ensure accessibility of the visual materials to people with color vision deficiencies. To further support inclusivity, we will highlight tools that can translate our workflows into other languages. Using GitHub's features, we will illustrate the benefits of receiving community contributions to projects. By coordinating with NASA's Exoplanet Program Analysis Group (ExoPAG) and the NASA Exoplanet Science Institute (NExScI), we will ensure that the benefits of ExoCore reach a broad audience in the field. We will collect anonymized feedback to construct

quantitative metrics to measure the impact of open science and attend the annual TOPS coordination meeting in D.C. to present our work and user feedback.

Our aim with ExoCore is to advance open science literacy, lower the entry barrier to exoplanet research by making it easier for new researchers to produce high-impact results, make major results and methodologies easier to find, accelerate the adoption of open science practices and principles, and reveal the discovery potential of public tools and data. ExoCore's focus on exoplanets supports science consistent with the funding priorities of SMD Divisions of Planetary Science and Astrophysics. The intrinsic merit of our proposal will be reflected in its ability to diversify and democratize exoplanet research in order to support the Open-Source Science Initiative and celebrate 2023 as the year of open science. Overall, ExoCore will facilitate NASA's coordinated efforts in the TOPS project, ensure accessibility and equity in research, and accelerate discovery in the exoplanet field.

Arthur Endsley/University of Montana, Missoula
Satellite Observations and Models Informing Agriculture: Training for Open Science Under Climate Change
22-TOPST22-0008

Projections of global climate change indicate that many agricultural regions will be warmer and drier. Elsewhere, intensification of the hydrological cycle may mean that rainfall is too intense, or less predictable, threatening established crop cycles and crop production. The study of these changes in the earth system is one of the goals of NASA's Science Mission Directorate (SMD), but the increasing reliance on computer-aided inference, including computational models and big datasets, requires skills that are neither thoroughly nor equitably distributed among domain scientists. Already, early-career climate scientists are spending too much professional time trying to solve frustrating issues of data access and software compatibility.

Effective, evidence-based decision making in agriculture and agronomy under a changing climate will require input from scientists that can combine models with large and diverse datasets, collaborating across open digital infrastructures. The reality is that the vast majority of domain scientists are self-taught when it comes to computer-aided inference and digital workflows. They lack the habits and best practices that are essential to identifying, acquiring, and analyzing NASA SMD datasets while ensuring reproducible and open results. This is especially problematic in less industrialized nations, where digital infrastructure is underdeveloped and training opportunities scarce.

We are a group of climate scientists, agronomists, and ecologists that have first-hand experience both teaching and learning how to use the tools and techniques of Open Science. We propose to develop a curriculum that will prepare our colleagues and students to conduct Open Science on climate and agricultural systems, with a focus on the essential elements of food security in semi-arid regions: Climate, Soil Health, Water Availability, and Crops. Our curriculum will cover the fundamentals of how climate datasets are produced, including remote sensing, General Circulation Models, and Earth System Models. Then, we will introduce each topic area (Climate, Soils, Water, Crops) by exploring how NASA SMD data, especially cloud-ready datasets, can be used to answer specific, applied questions related to food security in the Middle East and Northern Africa (MENA) region. Specifically, NASA's investments including the Earth Observing System, Soil Moisture Active Passive (SMAP), the Gravity Recovery and Climate Experiment (GRACE), and ECOSTRESS will all feature prominently as resources for insight into agricultural water availability, soil health, crop condition, crop stage, and production.

The five (5) ScienceCore modules we will develop as part of our curriculum will be based entirely in open-source software, implemented in Python, and distributed using literate programming tools like Jupyter Notebooks. We will make connections with NASA's OpenCore, filling in what we perceive to be gaps that learners frequently struggle with when first learning a programming workflow: how to manage large and diverse datasets; developing an explicable and reproducible workflow; separating

algorithmic code from runtime parameters; documenting code; and how to verify algorithms or procedures.

The MENA region shares similarities with many semi-arid food-producing regions of the world, including the western United States, and these regions will also feature as case studies. Open Science means that science should be accessible, legible, and freely available irrespective of geography or location; therefore, our proposal team, which includes scientists from the MENA region, will translate lesson materials into French and Arabic to make sure this curriculum is ready for immediate adoption in the Francophone Mediterranean, Sub-Saharan Africa, and the MENA region. Our collaborators in the MENA region will also assist in the teaching and dissemination of the curriculum to students and researchers at institutions in Algeria and Egypt.

Nicholas Halper/Neuromatch, Inc.
An Open, Community Supported, Accessible Summer School for Climate Science
22-TOPST22-0030

Open science promises faster innovation, increased rigor and accuracy, fewer wasted resources, and more equitable participation in scientific research by those from underrepresented backgrounds. Despite these benefits, a transition to open science practices is impeded by a lack of education in open science tools and principles, a lack of good examples to follow, and an oversimplification that open science is simply about data sharing.

To address these, we structure this proposal around three insights. First, that open science adoption is primarily an issue of cultural adoption and requires buy-in and participation of many stakeholders from students to investigators to educators. Second, to drive cultural change, open science must be both formally taught and demonstrated by example through high quality, engaging and inspiring materials that are built on open science principles themselves and open for use by many professors for adoption into their classrooms. Third, it is best to meet scientists where they are at and incorporate open science practices into domain-specific content where people are already learning.

Based on these insights, we will develop a summer school and curricula that merges NASAs OpenCore modules with high quality domain specific education for open science tools, data sets, and practices. Instead of developing a summer school that is only available while NASA funding is actively supporting it, though, we propose a focus on building a community around open science practice. We will develop an open educational program that is built on bidirectional communication between educators and learners and that utilizes hundreds of people contributing short/micro lectures and interactive lesson plans in jupyter books. The program features learning by doing, freely usable programming languages, live instruction, flipped classroom group based learning, and we will deliver it in several different languages and all online to ensure maximum accessibility, equity, and adoption. This initiative will be driven by a self-sustainable

funding model to ensure that it provides these benefits continuously and long term while being constantly updated through its open source nature. To support this continuous evolution, we will continue to meet scientists where and how they communicate by engaging with them at other workshops and conferences. We are not alone in this approach, and it has been demonstrated in other scientific fields, such as neuroscience [1, 2, 3, 4].

Through algorithmic networking mechanisms, the group classroom format, and built-in mentoring on student-lead research projects, we believe we can form both the tight and broad ties that students need to be truly integrated into the open science and climate science community while ensuring accessibility for all to the best education available on both climate science and open science topics.

Lisa Lowe/North Carolina State University
Building a framework for ScienceCore Carpentry from a Marine Sciences Lab
22-TOPST22-0007

Many research articles conclude by saying, this methodology may be used in other systems". The articles may be published in open access journals along with data and code, but does that equate to open-source science? Open-source science is transparent, accessible, reproducible, and inclusive. For any given article, how long would it take not only to reproduce the results but to use that knowledge to make new discoveries? The reader must fully understand the science and methodology, acquire the data, run the code, and complete the analysis.

In an undergraduate lab, students analyze a scientific topic then make calculations. The science is explained clearly, and computational tasks are simplified by using small, prepared datasets and a limited but fundamental set of coding techniques. If a researcher can present the science behind their article in the context of a marine sciences lab, then the science is accessible.

The lab used a small, prepared dataset, but the original data is required to reproduce the article. Getting satellite data can be a challenge, and the process used to prepare the data is rarely documented. The lab simplifies coding tasks, but the full analysis is complex, using custom code that ran on local resources within a custom environment. How can anyone reproduce those results, let alone use the techniques for other systems?

We will retrace the steps that would be needed to reproduce the results from two articles, one in physical oceanography and one in biogeochemistry. In the process, we will create an open-source computational component to a marine sciences lab, an interactive, online course that lowers the learning curve for using satellite data, and a recipe for transforming a traditional workflow into an open one. The recipe will be tested by issuing a challenge to reproduce the workflow, adapt the workflow for other systems, discover new science, and publish the results.

While the products created will be research specific, our methodology may be employed for any research topic in any discipline.

Kytt MacManus/Columbia University
Science Core Heuristics for Open Science Outcomes in Learning (SCHOOL)
22-TOPST22-0024

The project will develop curriculum in support of the Transforming to OPen Science (TOPS) project goals to increase understanding and adoption of open science principles and techniques, accelerate major scientific discoveries, and broaden participation in SMD-funded research by historically under-represented communities. The proposal is in alignment with the SMD's Science Plan, Explore Science 2020-2024: A Vision for Scientific Excellence, the project will specifically advance STRATEGY 4.2: Purposefully and actively engage with audiences and learners of all ages to share the story of NASA's integrated science program with a specific focus on undergraduates, graduates, and professional teams. The proposal advances the NASA 2022 Strategic plan Objective 4.1: Attract and develop a talented and diverse workforce.

We will extend the approach developed by OpenCore with its emphasis on student understanding of open science values, tools and communities, and applications in order to teach about the data science life cycle. This will be accomplished by creating open, interactive, interdisciplinary content on Earth Science Applications use cases. We will develop lessons across the thematic areas of water resources, agriculture, health and air quality, disasters, climate, wildfires, and environmental justice, integrating within each the additional thematic areas of population and infrastructure. Within lessons we will demonstrate how to access and analyze NASA data sources in open data science lifecycle workflows which draw heavily on data and services from NASA Distributed Active Archive Centers (DAACs). We will utilize core open-source data, analysis, and visualization libraries such as GDAL, Numpy, GeoPandas, rasterio, exactextractr, duplicatoR, pesR, matplotlib, ggplot, and others for the creation, management, and sharing of reproducible science. We will adopt an inclusive teaching style with an emphasis on active learning to guide users through generation, collection, processing, storage, management, analysis, visualization, and interpretation of NASA data.

We will employ a Scaled Agile Framework (SAFe) that will use free and open technologies (Google Tables, GitHub, Jupyter Books) to implement SAFe best practices for content creation management. We will form agile teams consisting of Subject Matter Experts (SMEs), developers, and a diverse set of end users from various ethnic, linguistic, and socioeconomic groups. We will offer small stipends to students in historically underrepresented groups to participate in the development iterations and test materials. Over the course of the 2 year project we will develop and release modules at regular intervals by conducting a series of sprints devoted to specific thematic areas in hybrid open meetings. These meetings will help to engage diverse user communities to

determine module needs and use-cases: for example, the well-established Population-Environment Research Network (PERN) and State University of New York Educational Opportunity Program (SUNY EOP), or the in-development Center for Aging, Health and Climate Change. The content we develop will be suitable for both self-driven, or instructor-led learning, and will be available in both English and Spanish translations.

Robyn Moncrief/National-Louis University
Ensuring Culturally Responsive Practices and Community Building in Open Science
22-TOPST22-0021

Ensuring Culturally Responsive Practices and Community Building in Open Science (Project CRCB) applies innovative approaches to increase the success of diverse undergraduates and science community members while engaging with open science. The need for efforts to make science more accessible is clear. The rates at which Black and Latinx students successfully pursue science at the undergraduate level and into careers are low, and the workforce's diversity varies widely across jobs and fields. Project CRCB directly builds on findings from the Institute for Higher Education Policy report on NSF's Model Replication Institutions Program, Diversifying the STEM Pipeline: The Model Replication Institutions Program (Cullinane, 2009). In considering lessons learned and the replicability of best practices, the report spoke to the deep need of minority students to connect to academic and professional communities; and find connection between their academic work and community enhancement. Project CRCB will create open science summer school opportunities that use community-building activities and culturally responsive teaching practices to provide a more accessible learning experience. This project will leverage National Louis University's growing STEM community and efforts to provide an accessible learning environment through culturally responsive teaching practices. Project CRCB will organize a series of 5-day summer school sessions, engaging undergraduates and science community members through the completion of the five OpenCore modules. These summer school sessions will also include networking and collaboration opportunities, approaches to bringing OpenCore curriculum back to the attendees' own communities, techniques in teaching OpenCore with culturally responsive teaching practices, and opportunities for culturally responsive science mentorship within local communities.

Madicken Munk/University Of Illinois, Urbana-Champaign
Bringing Together Open Science and Research Software
22-TOPST22-0023

Bringing Together Open Science and Research Software
Open, reproducible, and accessible science builds community trust, facilitates confidence in published data, and creates accessible pathways for science education. Open science

also depends on the reliability of infrastructure to ensure analyses pipelines reliably reproduce results. Research software is a key critical component of this open science infrastructure. Ensuring that research software is stable, well-maintained, and accessible will strengthen outputs in open science. This will contribute to the long-term success of open science deliverables.

Research software is used widely in the sciences, engineering, and the social and behavioral sciences. Though research software is critical infrastructure for science, it is often maintained by a small group of consistent developers. With limited resources for maintenance, development, and user support, software can stagnate and a new project will be written, built, and disseminated to replace it. Users are forced to adapt and shift, wasting precious time and resources, and the cycle continues. Many efforts already exist to create pathways and onboard new research software developers to address this problem from the top-down". However, few efforts center the contributions of a vibrant user community to sustainable research software. Users are key in finding code corner-cases, in discovering bugs and raising issues, and in continuous evaluation of code reliability. However, there is often a gap in the perspective and needs of users and developers. In this work we propose to develop four modules of training material on how to navigate a research software community as a user, and the best practices for crediting, engaging with, and working with the research software development community. This curriculum will help nurture a positive user community culture for research software, which in turn will contribute back to the research software community.

We will use these research software training modules in concert with the TOPST curriculum on principles of Open Science to educate and train researchers through a series of 6 summer schools with 60 learners each. Summer schools will use inclusive methods of instruction and have earmarked funding for accessibility needs to facilitate inclusion of learners of varying needs. Learners from these schools will be fluent in navigating research software for their research and will be able to generate high-quality open science deliverables. They will be able to meaningfully contribute to the research software community as a user of the software, with open science publications that cite and credit software. They will be able to help other users navigate issues, and they will help foster a positive community in their research software networks. Together these curricula will help researchers contribute to both sustainable science and sustainable software.

James Munroe/Code For Science And Society, Inc.
Reproducibly Analyzing Wildfire, Drought, and Flood Risk with NASA Earthdata Cloud
22-TOPST22-0014

As the climate changes, prediction and management of the risk of wildfire, drought, and floods has become increasingly challenging. It is no longer sufficient to assume that what has been normal and historic for the last century will occur with the same frequency into

the future. These natural risks are intrinsically linked to the changing distributions of surface water, precipitation, vegetation, and land use in both time and space. At the same time, there are now hundreds of petabytes of relevant Earth science data available through the NASA Earthdata Cloud that can be used to understand and forecast these water-dependent environmental risks. With the volume of Earth science data growing dramatically year over year, it is important for scientists to understand how open science and cloud-based data intensive computing can be used to reproducibly analyze and assess the changing risk profile of wildfire, drought, and floods.

In this proposed TOPS ScienceCore module, learners will learn to identify, extract, analyze, visualize, and report on data available through NASA Earthdata Cloud for three different scenarios: wildfire, drought, and flood risk. The module will build upon TOPS OpenCore and reinforce principles of reproducibility and open science-based workflows. Computationally, the scenarios will estimate changes in the hydrological water mass balance for defined regions primarily using remote sensing data. We will demonstrate best practices in data-proximate computing" by considering examples that involve computing climatologies and other statistics from long-time series using numerical methods that scale well with the data being available on the cloud. This module will leverage scientific Python libraries such as Xarray and Dask to perform the computations. The focus of this module will be on data processing and visualization and doing so in a reproducible and transparent way.

After completing this module, learners will be able to adapt and remix the scenarios for their own open science objectives regarding environmental risks such as wildfire, drought, and flood. These risks are common worldwide yet need to be each analyzed in a regional context. The module will provide concrete examples that showcase how open science can be done.

The module will be written as an extension to the OpenCore framework and all course materials will be open, available in English and Spanish, and accessible in the vision, hearing, mobility, and attention dimensions. The ScienceCore module will be released as one or more Jupyter notebooks on GitHub with supporting material for delivering the course using the cloud either for in-person or for virtual cohorts.

Cameron Riddell/Dontusethiscode LLC
Virtual Cohorts: Developing Lifelong Committed Interaction With Open Science
22-TOPST22-0029

The Year of Open Science (2023) is upon us.[1] We face an unprecedented opportunity to advance the mission of Open Science through NASA's Transform to Open Science initiative and its OpenCore curricula.
The NASA Transform to Open Science mission has developed training to educate scientific researchers on foundational topics in Open Science. The OpenCore curricula will be offered as an online course, the goal of which is to spark change and inspire open

science engagement& that will shift the current paradigm."[1] The TOPS program is only the starting point,"[1] and its success at developing and supporting foundational engagement is key to the success of the overall mission.

This proposal outlines a strategy for guiding scientists not only to complete the OpenCore curricula but also to develop a career-long engagement with Open Science. This strategy includes ongoing community support to help members incorporate Open Science ideas into all aspects of their work.

Studies of online courses have uncovered that an important predictor of success in large online courses lies in their ability to establish a sense of a learning community"[4,5]. Moreover, large online courses suffer from high attrition rates[7,8]. Tools used to create such learning communities include interactive online forums, instructor-led Q&A sessions, and external social media groups[4]. By participating in a community developed around an online course, learners have the opportunity to enter into their own discussions, ask and receive clarity on lingering questions, and, importantly, network and motivate one another towards course completion[5].

We propose the creation of a learning community around the NASA TOPS OpenCore curricula. In this community, course participants will be given the opportunity to attend regularly scheduled workshops which highlight practical aspects of the course content, to hear from prominent guest speakers discussing how they have applied Open Science in their research, and to interact with each other, offering encouragement and peer-based expertise.

By creating a community centered around the NASA TOPS online course, we will positively impact course completion rates of the OpenCore curricula. Beyond this goal of increased course completion, we will develop a strong community centered around the TOPS mission, composed of individuals who have completed the online course and intend to incorporate Open Science methodologies into their research.

Lauren Sanders/NASA Ames Research Center
Training in Artificial Intelligence and Machine Learning for Space Biological Sciences Using NASA Cloud-Based Data
22-TOPST22-0020

Here we propose to develop Open Science training modules for artificial intelligence and machine learning (AI/ML) in space biosciences. We will design modules in the OpenCore style and format, in 2.5 hour interactive learning experiences to be taught in the mornings during a week-long meeting. Each week-long training course will cover a different topic, including fundamentals of AI/ML and space bioscience data, ethics and best practices for AI/ML and Open Science, and use cases which train students in specific algorithms to solve real-world problems using cloud-based NASA data.

Donald Winston/Polyneme LLC
Heliophysics ScienceCore Curriculum Development with Emphasis on Knowledge Representation Techniques to Increase Usability of NASA Cloud-Based Datasets.
22-TOPST22-0025

The key, central objectives of this proposal are to

1. Develop showcase notebooks (i.e., Jupyter notebooks) around cloud-based NASA heliophysics and space weather datasets that engender a sense of excitement about open science (methods and outcomes) within the general science community by demonstrating powerful, interactive infrastructure for querying, visualization, investigating data governance and lineage, data lake navigation, and dataset discoverability.

2. Bootstrap a knowledge graph (KG) via coordinated human-expert modeling of taxonomies/ontologies and curation of high-quality metadata that describe said datasets.

3. Advance appreciation among domain specialists and nonspecialists of semantic knowledge representation approaches that lower data preparation efforts, enable deeper analyses based on enriching data context, and facilitate gathering a critical mass of domain awareness by interpreting and interlinking data from different sources.

Critical methods and techniques proposed to accomplish the stated objectives are:

1. Literate programming, as realized by the Jupyter project and, in particular, the Jupyter notebook format and the iPython kernel for interactive computing through a web-browser-based interface, as well as mature infrastructure for deploying environments for user sessions such as BinderHub for JupyterHub containers.

2. Resource-oriented (e.g., RESTful) HTTP APIs to query and subset cloud-based datasets for interactive exploration via e.g., Jupyter notebooks.

3. W3C Semantic Web standards for machine-actionable (FAIR) knowledge representation based on Web technologies. In particular, the Resource Description Framework (RDF) suite of standards for metadata querying (SPARQL), serialization (e.g., Turtle), exchange (e.g., HTTP media types), modeling and inferencing (e.g., RDF entailment), and validation (SHACL), as well as tooling (e.g., the Python RDFLib project) and infrastructure (e.g., open-source RDF graph databases, process-embedded or otherwise).

The perceived significance of the proposed work to the objectives of the solicitation and to NASA interests and programs in general:

1. The bootstrapped knowledge graph will empower computer systems that in many aspects surpass the analytical capabilities that a typical human specialist has in their area

of study. And unlike neural network models, the KG is human readable and explainable one can use it as a reference data structure, correct it, govern it, publish it, etc. This is expected to allow more automation in scientific data management and to lower preparation efforts needed for AI projects across NASA that pertain to heliophysics and space weather.

2. By demonstrating the application of machine-actionable knowledge representation via showcase notebooks, the utility of semantic formalisms will be more accessible and concrete to both heliophysics / space weather specialists and to a broader community of nonspecialists that may interoperably interface with this field's open datasets.

Yo Yehudi/Code For Science And Society, Inc.
Teaching TOPS OpenCore by Embedding Community Values.
22-TOPST22-0015

Objectives:

We propose to deliver fully virtual training cohorts for the NASA TOPS OpenCore training curriculum with a highly qualitative component throughout the training. We propose five 6-week Virtual Cohorts per year, for 3 years, that will include guided reflection and hands-on implementation of open science concepts, processes, and techniques. This program will teach one OpenCore Module per week, by bringing in expert speakers from across different scientific domains, allowing learners to actively embed their training in their work and reflect on their learning with a mentor.

Methods:

The proposed training will require approximately 4 hours of personal commitment per week. This includes a mix of online training from domain experts to cover the 5 TOPS OpenCore modules, discussion groups to explore nuance, hands-on work applying open principles, methods, and techniques to a trainee's work, and 1:1 mentoring. Pedagogy training will ensure that people who participate in the training become ambassadors and multipliers", equipped with the right capabilities to pass on their knowledge and skills. Mentors have hands-on experience as open science practitioners, and are provided professional coaching training to ensure that participants are effectively guided through their open science journey.
Additionally, community immersion will be facilitated via Slack, where cohort participants, mentors, and experts will be able to interact before, during, and after the 6-week program.

Significance of the proposed work

This proposal advances open science literacy among a wide and diverse scientific community, including people from historically underrepresented and marginalized

communities. The proposed mentoring and training aims to allow open and equitable communities to disseminate their work, as it is designed to fully engage and embrace a multiplicity of talents, ideas and perspectives. Inclusion by design is one of the core actionable principles we strive to follow and advocate for. We see open science literacy as deep understanding and adoption of approaches and techniques but also of principles and thus aligns closely with NASA TOPS objectives and priorities.
Our communities welcome a large number and wide variety of scientists and researchers, from undergraduate students to senior practitioners. In this context, we make explicit efforts to center an ethics of care, encouragement and support in the community that benefits everyone, newcomers but also experienced scientists. An example of such commitment is the microgrants, which have been provided to cohort participants to support and encourage their engagement in the community
The principle of mutual help is centerstage. We are ready to provide scientists with the help, encouragement, and equal access needed for completing the OpenCore curriculum, earning their badge and making the Virtual Cohorts a success.

This application complements another proposal, but is English-language first.
