1. Project Summary

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.

2. Scientific/Technical Management

2.1. Introduction

Culture change in academia is hard, and many initiatives to change science culture come from grassroots organizations created by academics who wish to see research culture as open and equitable [1], [2]. Grassroots efforts are only part of the equation; the other part is support from funders, policymakers, and journals. We propose to train scientists, researchers, and research-supporting staff at all career stages. This training will push academic culture away from seeing open science as an important "extra", ultimately skipped because it takes time and does not reap rewards, towards seeing open science as essential [3]–[6]. By embedding open science practices into scientists' daily work, our goal is for them to move towards **responsible open sharing as the bare minimum** they expect from themselves and others, similar to the way science views papers and peer review: an essential part of the scientific process.

2.2. Objectives and Expected Significance

Objective 1: Train at least 5 cohorts of ~50 people per year for 3 years (750 to 1,050 participants). Each cohort will receive 6 weeks of online training covering the 5 NASA Transform to Open Science (TOPS) OpenCore modules in Spanish.

Objective 2: Steward a supportive and inclusive Spanish-speaking online community around NASA TOPS training. Offering the TOPS OpenCore in Spanish is essential for lowering the language barrier and, therefore allowing the participation and community engagement of Spanish-speaking scientists.

Objective 3: Skills are retained, re-used, and shared onwards. Rather than short-form training devoid of a strong community around it, the proposed 6-week Virtual Cohorts provide opportunity for learning and hands-on implementation of open science. Community-based training allows for continued engagement beyond the 6 weeks of the program.

Objective 4: Create a "multiplier" effect. In order to enable training and the building of an inclusive open science community, actions taken in Objectives 1, 2, and 3 will be further facilitated by pedagogical and movement building training for participants and mentors, so that even after they have completed the course, they continue to train others and embed good practices in their labs and institutions. This is elaborated further in the impact section.

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

2.3. Impact of proposed work

Our training is designed to embed practical, attainable real-world skills into daily behavior, rather than teaching participants about best practices that are difficult to achieve. By offering participants "quick wins" that are easy to implement, alongside ways to easily share their new-won knowledge, we intend to create **open science ambassadors** who spread open science skills to colleagues and collaborators, including those who have not yet participated in a cohort or completed OpenCore badges.

This is achieved not only by participants completing the training, but also by participants working throughout the 6 weeks of a virtual cohort in a project of their interest, with time for weekly reflection while being guided by a professional mentor who can offer 1 on 1 support to achieve the cohort program objectives. Finally, all engaged participants who are likely to be capable mentors are encouraged to participate in the program in future rounds as mentors, helping to embed these skills even more deeply, and passing on their new-found skills.

By focusing significant effort on marginalized and lower-resource communities, we hope to not only act as Allies (allies are people with more privilege or power in a situation who support and elevate others with less privilege – often from marginalized groups or identities), **but also as role models for how equitable collaborations in open science can be created, without exclusion, tokenization, exploitation, or extractivism.** [7], [8]

2.4. Relevance of proposed work to announcement

TOPS Goal: "Increasing understanding and adoption of open science." We will advance open science literacy among a wide scientific community. Our proposal has a 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. Open science literacy means to us deep understanding and adoption of approaches and techniques, but also of principles.

TOPS Goal: "Accelerate major scientific discoveries." Open science accelerates scientific discoveries and innovation [9]–[12]. By training and stewarding a community of between 750 and 1,050 new open science ambassadors, we will increase open science literacy, which will serve the final goal of accelerating scientific discoveries.

TOPS Goal: "Broaden the participation in SMD-funded research by historically

under-represented communities." Inclusion by design is one of the core actionable principles we advocate for and implement in all our work. Many researchers struggle to find right avenues to learn and reflect on taught skills in the context of their own work, which our personalized mentoring approaches effectively address. Virtual cohorts with specific accessibility measures make it equitable for people with a broad range ability and disability to engage with training in open science. This proposal puts Spanish first, thus, Spanish-speakers, who are numerous in the US, will have access to virtual cohorts that lower the language barrier. In addition, virtual training in Spanish will make it possible to reach other global communities who speak Spanish and have an interest in open science.

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

2.5. Approach and methodology

Open-science-curious researchers, from undergraduates to senior practitioners, will be asked to commit 4 hours a week for 6 weeks. Learning materials, synchronous Zoom calls, and all interactions will be in Spanish and online with inclusive and accessibility measures. This will be a mix of online training 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.

All cohort facilitators, mentors, and experts will be offered training on coaching and pedagogy for delivering online training. When needed, domain experts will be offered professional-quality live-interpretation from English to Spanish. Mentors will be experienced open science practitioners, with professional coaching training. Asynchronous interactions and community engagement will happen via a Spanish-speaking Slack workspace where cohort participants, mentors, experts, organizers, and other open science practitioners will be able to interact before, during, and after the 6-week program.

2.5.1. Detailed plan for the virtual cohorts

Size	Average of **50 participants per cohort**, formed of self-selected team projects and individuals. Capped at 70 participants total.
Location	Zoom-based online collaboration, Slack, mailing lists.
Duration	**6 weeks** per cohort, 5 cohorts per year.
Scheduling	Up to **4 hours per week** for participants, including 2.5 hours of cohort call, a 30 minute 1:1 mentor discussion, and reflection time.
Cost	Individuals $700, Groups of 2-5: $1,200. Equity access: no-quibble fee waivers for anyone who asks, and microgrants to support participation.

Based on historic data from prior open science training cohorts that have been run [13], 50 participants per cohort is an appropriate number for the emergence of a cohesive and active community. Cohorts will be a mix of self-selected groups and individual participants. We will work together with NASA and other virtual cohort training offerings to agree on ways to effectively reach participants.

Cohort timeline		Before the course			Week					
		8 weeks	4 weeks	2 weeks	1	2	3	4	5	6
Pre-cohort	Application period opens	x								
	Application period closes		x							
	Applicants notified			x						
TOPS OpenCore module 2.5 hours each	**Ethos**				2:30					
	Data					2:30				
	Software						2:30			
	Results							2:30		
	Tools & Resources								2:30	
Final Graduation	Group live-streamed presentations									2:30
Mentor (30 mins)	1:1 or 1 mentor:group project				0:30	0:30	0:30	0:30	0:30	
Reflection & implementation	Self-led by participants - optional assignments - prep for final presentations				1:00	1:00	1:00	1:00	1:00	1:30
	Weekly hours				4:00	4:00	4:00	4:00	4:00	4:00

Figure 1: visualization of participant time commitments during the virtual cohort

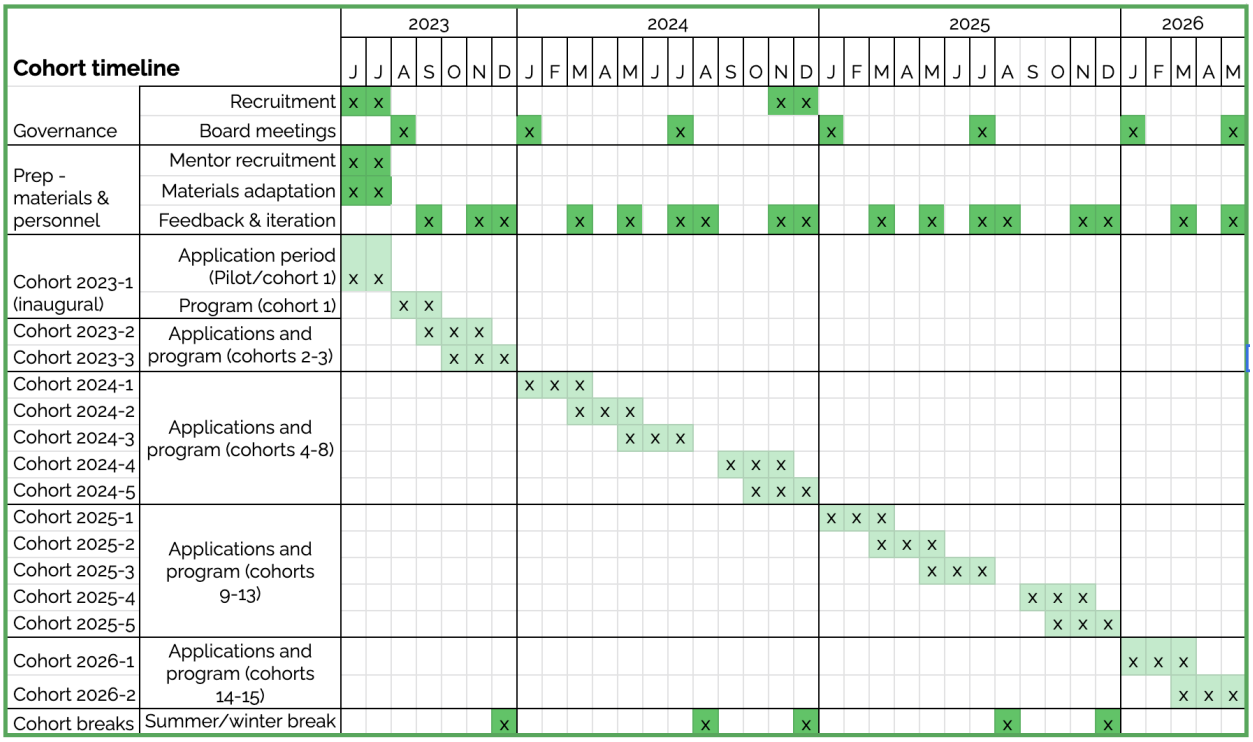

Figure 2: GANTT chart showing anticipated timings of all fifteen cohorts.

Application period: Requiring a short written application will allow the cohort organizers to more effectively assign mentors and/or create “themed” cohorts with common interests if any individual cohort is oversubscribed. It will also allow us to make sure all feasible accessibility accommodations are provided. Each group or participant will be asked to set a goal or project when applying, in order to get hands-on experience applying the principles they learn.

Location and time requirements: Participants will commit approximately 4 hours to this course per week. This includes a mix of online Zoom-based training from domain experts to cover the 5 TOPS OpenCore modules (2.5 hours a week), discussion groups to explore nuance, methods, and techniques, and 1:1 mentoring (30 minutes per week). Mentors will be asked to commit 1 hour per week to each participant or group (30 minutes call + prep time), and generally will mentor two to three projects or individuals.

Cost and Fee waivers: We would prefer to keep the cohorts 100% free to all attendees, as we believe that the most equitable training is accessible to all. In order to ensure we have sufficient budget to also pay our mentors, call hosts, and other community participants an honorarium for their services, however, we have instead planned for approximately 50% of each cohort to pay for their training, and up to 50% will be offered no-quibble fee waivers upon request. As such, fee-paying groups of 2-5 will be able to participate for approximately $1,200 per group and individuals will pay $700 per

individual. This fee structure seeks to incentivize participants to apply in small groups, as we have seen researchers learn about open ways or working best when practicing them in peer-to-peer interaction [14], facilitated within small groups during their training.

We will closely monitor this fee structure and adjust it if needed to ensure we get a diverse and representative group of participants across protected classes such as race, sexuality, disability, and gender.

Participants will be welcome to join even if they have already completed badges for one or more of the individual modules. They will be permitted to skip weekly calls for badges they have completed, but we will still encourage them to join and participate in order to form network ties and participate in breakout groups with others in their cohort.

Feedback and iteration: As shown in Figure 2, we will start by preparing cohort logistics and setting out expectations for the first virtual cohort. This includes training mentors and experts in the specifics of translated and localized OpenCore materials, forming template emails, meeting notes, application forms, and information for cohort members (including participants, mentors, and trainers) so they can have a clear understanding of the time commitments required. After these preparations we will run a first cohort that will be a pilot for the next 14 virtual cohorts.

Concurrently with the setup of the inaugural cohort, we will recruit mentors, trainers, and experts from our extended open science networks, and particularly inviting applications from Spanish-speaking participants who authored the original OpenCore material [15].

Project governance will be driven not only by the PI and Co-Is, but also directly from the community. As a group, we will create team working principles, governance, and conduct expectations by community consensus, including a framework for conflict resolution, and how decisions are made, as a community rather than top-down. To this end, we will adapt the processes used successfully by other community-based organizations to design their governance [16]. This will allow us to avoid inadvertent tokenization, and intentionally build a healthy Spanish-speaking community around the materials and subject matter, including culturally relevant nuanced localisation and community-driven values.

2.5.2. Risks and mitigations

Risk: Insufficient Spanish-speaking virtual cohort applicants.

Mitigation: This risk is unlikely because, in addition to our current reach to Spanish-speaking researchers worldwide, we started conversations with two US-based Hispanic Serving Institutions (HSI) to ensure we will have deep reach to everyone who

may benefit from an OpenCore virtual cohort fully in Spanish. Nonetheless, in the unlikely situation where there are insufficient Spanish-speaking virtual cohort applicants to meet our goals, we are bilingual and will reconvert to English, complementing our sister English-centered proposal.

Risk: By offering Spanish-speaking virtual cohorts, Spanish-speaking researchers might mingle only with each other and be isolated from the English-speaking open science community.

Mitigation: It is often believed that since English is the scientific lingua franca, every scientist should learn to speak it. In that way, there would not be a need to have virtual cohorts in Spanish. In our experience, this is a recipe for exclusion. Allowing Spanish-speaking scientists to learn in their first language, equalizes their open science knowledge to international standards. By teaching open science lingo and practices first, it becomes significantly easier to join the international open science community, without first requiring that all open science-curious researchers learn an entirely new language. Also, we have experience running multilingual communities of practice, including channels for Spanish-only or English-only exchanges where bilingual persons act as liaisons between the two groups that end up benefiting from each other in spite of the language barrier. Automated services such as Google Translate offer additional "safety cushions" that enable people to commingle with only minor translation errors.

Risk: Insufficient Spanish-speaking open science experts to teach the OpenCore modules.

Mitigation: Invite English-speaking open science experts and run events with high-quality bilingual interpretation, which we have done several times already.

Risk: OpenCore modules are not still finalized. It is risky to offer training for a material that is unknown.

Mitigation: Part of our team has access to and has participated in the development of OpenCore materials. On October 11th, Dr. Gentemann, TOPS Program Scientist, praised these materials using the following words "we [referring to NASA] are really happy with all that [material]. Everytime I read it, I feel like I learn something new because it brings this global perspective and deepens my experience about open science. I think it is absolutely wonderful" (Quoted at 1:50:28 in [17]). In the interim, we are relying on outputs from OpenSciency, as deposited on Zenodo [18], for our expectations of what the materials might look like.

Risk: Insufficient time and funds to translate OpenCore modules to Spanish.
Mitigation: We will run a community-driven "NASA TOPS OpenCore Modules Spanish Translatathon" between April 2023 (TOPS materials release date) and May 2023, with

the goal of finishing translating OpenCore Modules before this proposal starts. We are in the process of seeking funding for this event, for instance, using the NASA ROSES Topical workshops, symposia, and conferences (TWSC) funding mechanism. Additionally, one of our staff members has recently been awarded a Fellowship to sponsor them creating translations of open science training materials.

Risk: Insufficient Spanish-speaking open science experts with pedagogy for virtual teaching skills.

Mitigation: We have access to a community of 48 Carpentries-certified [19] Spanish-speaking instructors. We are also organizing the workshop "Training Spanish-Speaking Trainers for Teaching NASA TOPS OpenCore Modules in Spanish" between May and June 2023 (this proposal's start). We are in the process of seeking funding for this event, for instance, using the NASA ROSES Topical workshops, symposia, and conferences (TWSC) funding mechanism.

Risk: Insufficient Spanish-speaking mentors recruited.

Mitigation: Mentors are usually community members external to our organizations with experience in open science practice. To help mentors prioritize their workload, we offer honorarium payments. If needed, staff of the proposing organization can perform the role of mentor internally. After the inaugural round of training, however, we will recruit more mentors from the previously graduated cohorts, so we anticipate this would not be a longstanding challenge. Local recruiting in Latin America will also be a key element of the fellowship awarded to one of the Co-Is to work on building scientific and technical capacities through the co-creation of networks, learning spaces, and accessible resources for Spanish-speaking communities. [20]

2.6. Management Structure

Inclusion is at the core of our working practices, so much of our plan for coordination and governance revolves around the community, taking a transparency-first and community-driven approach. This proposal is created by two teams who have been regularly working hand-in-hand since early 2020, and have shared visions and values. We will begin by creating temporary team governance structures for the kickoff of the project, but as part of our project ramp-up, we will ensure we set up a team charter, governance structure, and decision making process, co-created with stakeholders, including NASA (where appropriate), and researchers representing various different career stages and backgrounds.

This governance system will be specifically designed to act as a system of checks and balances on our openness, our pedagogy, and our inclusion. We regard this open community governance as an essential part of embodying the behaviors we wish to

teach around the benefits of open collaborative work. These governance policies and procedures will be online and open from the point of initial development, even when in "rough" format, unless personal protection is needed, for example in Code of Conduct enforcement settings.

Open communities, however, do not run alone without people power. Our dedicated staffing team and Co-Is will coordinate the day-to-day running of the cohorts and the governance processes, using the same processes and procedures we intend to teach as part of TOPS OpenCore. Our planned reflection and iteration time will apply not only to course content, but also to management, processes, and governance.

2.7. Description of Contributions

PI will be responsible for: the operations planning to complete all deliverables in due time and managing the team, all hiring, expense authorization, reporting, achieving our equal access goals, and measuring the impact of the virtual cohorts.

Co-I 1 will work hand-in-hand with the PI leading cohort operations and coordinating strategic vision between leadership and management.

Co-I 2 will focus on the financial management of the cohorts, including microgrant participant support policies, accounting and financial reporting for cohort income, and data protection.

Co-I 3 will be in charge of providing any required infrastructure for the cohorts (alongside Co-I 4 and Co-I 5) by recruiting from and nurturing communities in parts of the US with a high proportion of Hispanic and Latino populations.

Co-I 4 will be in charge of running the cohorts together with Co-I 3 and Co-I 5, particularly working on all aspects related to the accessibility of virtual cohorts.

Co-I 5 will be responsible for coordinating cohorts. Responsibilities include managing applications, matching mentor/mentee pairs, arranging weekly expert speakers to deliver the OpenCore material, nurturing the community, and virtual cohort weekly communications.

The whole team will work together to measure the project's impact and openly share its results with the scientific community.

3. References

[1] M. Sharan *et al.*, 'Open Life Science (OLS) response to UNESCO global call for best practices in open science', Zenodo, Jul. 2022. doi: 10.5281/zenodo.6841873.

[2] K. Armeni *et al.*, 'Towards wide-scale adoption of open science practices: The role of open science communities', *Science and Public Policy*, vol. 48, no. 5, pp. 605–611, Oct. 2021, doi: 10.1093/scipol/scab039.

[3] Y. Kim and P. Zhang, 'Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories', *Library & Information Science Research*, vol. 37, no. 3, pp. 189–200, Jul. 2015, doi: 10.1016/j.lisr.2015.04.006.

[4] Y. Zhu, 'Open-access policy and data-sharing practice in UK academia', *Journal of Information Science*, vol. 46, no. 1, pp. 41–52, Feb. 2020, doi: 10.1177/0165551518823174.

[5] S. E. Ali-Khan, L. W. Harris, and E. R. Gold, 'Motivating participation in open science by examining researcher incentives', *eLife*, vol. 6, p. e29319, Oct. 2017, doi: 10.7554/eLife.29319.

[6] Y. Jeong and Mayiwar, 'The Dilemmas of Practicing Open Science', *BI Business School*, Jun. 10, 2022. https://www.bi.edu/about-bi/news/2022/06/the-dilemmas-of-practicing-open-science/ (accessed Dec. 05, 2022).

[7] K. Whitaker and O. Guest, '#bropenscience is broken science - The British Psychological Society', Sep. 20, 2020. https://www.bps.org.uk/psychologist/bropenscience-broken-science (accessed Dec. 06, 2022).

[8] D. Haelewaters, T. A. Hofmann, and A. L. Romero-Olivares, 'Ten simple rules for Global North researchers to stop perpetuating helicopter research in the Global South', *PLOS Computational Biology*, vol. 17, no. 8, p. e1009277, Aug. 2021, doi: 10.1371/journal.pcbi.1009277.

[9] Gentemann, Chelle and Ivey, Yvonne, 'NASA Tranform to Open Science', Nov. 2022, doi: 10.5281/ZENODO.7332651.

[10] J. Houghton, 'Open Access: What are the Economic Benefits? A Comparison of the United Kingdom, Netherlands and Denmark'. Rochester, NY, Jun. 22, 2009. doi: 10.2139/ssrn.1492578.

[11] S. Verhulst and A. Young, 'Open Data Impact: When demand and supply meet', GovLab, Omidyar Network, Mar. 2016. [Online]. Available: https://odimpact.org/files/open-data-impact-key-findings.pdf

[12] M. P. da Costa and F. C. L. Leite, 'Open access in the world and Latin America: A review since the Budapest Open Access Initiative', *Transinformação*, vol. 28, pp. 33–46, Apr. 2016, doi: 10.1590/2318-08892016002800003.

[13] Open Life Science, 'OLS Cohorts', *OLS Cohorts*. https://openlifesci.org/index#cohorts (accessed Nov. 28, 2022).

[14] G. Wilson, 'Teaching Online - Online Engagement', in *Teaching tech together: how to make lessons that work and build a teaching community around them*, Taylor & Francis, 2019. [Online]. Available: https://teachtogether.tech/en/index.html#
[15] NASA, 'TOPS Curricula Development Team', Nov. 29, 2022. https://github.com/nasa/Transform-to-Open-Science/blob/f30b23bcaca9e6e40be3cb696f4f24d112351a01/docs/Area2_Capacity_Sharing/OpenCore/OpenCore_leads.md (accessed Nov. 30, 2022).
[16] MetaDocencia Team, 'MetaDocencia's Governance', Dec. 2022, doi: 10.5281/ZENODO.7399003.
[17] *TOPS October 2022 Community Panel – Day 2*, (Oct. 11, 2022). Accessed: Dec. 01, 2022. [Online Video]. Available: https://www.youtube.com/watch?v=G5M7vThQu0s
[18] Almarzouq, Batool *et al.*, 'Opensciency - A core open science curriculum by and for the research community', Dec. 2022, doi: 10.5281/ZENODO.7392118.
[19] The Carpentries, 'The Carpentries', *The Carpentries*. https://carpentries.org/index.html (accessed Dec. 01, 2022).
[20] M. Black, '2023 SSI Fellowship Application - EspañOLS: empowering Open Science ambassadors in Spanish-speaking communities', Dec. 06, 2022. doi: 10.5281/zenodo.7405807.
[21] R. Joo *et al.*, 'Ten simple rules to host an inclusive conference', *PLoS Comput Biol*, vol. 18, no. 7, p. e1010164, Jul. 2022, doi: 10.1371/journal.pcbi.1010164.
[22] Y. Yehudi, K. S. Whitney, and M. Sharan, 'Enhancing the inclusivity and accessibility of your online calls'. OSF Preprints, Nov. 24, 2020. doi: 10.31219/osf.io/k3bfn.
[23] I. G. Poggio, P. A. Loto, M. Rajngewerc, A. Bellini, and Y. B. Saibene, 'MetaDocencia's Accessibility Profiles', Apr. 2022, doi: 10.5281/ZENODO.7387766.
[24] A. Bahji, L. Acion, A.-M. Laslett, and B. Adinoff, 'Exclusion of the non-English-speaking world from the scientific literature: Recommendations for change for addiction journals and publishers', *Nordic Studies on Alcohol and Drugs*, p. 14550725221102228, Jun. 2022, doi: 10.1177/14550725221102227.
[25] U.S. Census Bureau, 'S1601: LANGUAGE SPOKEN AT HOME - Census Bureau Table', *2016-2020 American Community Survey 5-Year Estimates*, 2020. https://data.census.gov/table?q=languages&tid=ACSST5Y2020.S1601 (accessed Nov. 28, 2022).
[26] Y. B. Saibene, P. Corrales, L. Ación, and N. Palopoli, 'Teaching to teach between communities in Spanish', Nov. 2021, doi: 10.5281/zenodo.7386613.
[27] W3C, 'Web Content Accessibility Guidelines (WCAG) 2.1'. https://www.w3.org/TR/WCAG21/ (accessed Dec. 06, 2022).
[28] AbilityNet, 'Accessibility Resources'. https://abilitynet.org.uk/accessibility-services/useful-resources (accessed Nov. 30, 2022).

[29] S. Zhang, K. H. Wapman, D. B. Larremore, and A. Clauset, 'Labor advantages drive the greater productivity of faculty at elite universities', *Science Advances*, vol. 8, no. 46, p. eabq7056, Nov. 2022, doi: 10.1126/sciadv.abq7056.
[30] L.-K. Huse, 'Fundamentals of Social Identity and Social Justice: Considering Social Identity Within CBPR With Marginalized Populations', *Frontiers in Communication*, vol. 5, 2020, Accessed: Dec. 06, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcomm.2020.00041
[31] R. F. Korte, 'A review of social identity theory with implications for training and development', *Journal of European Industrial Training*, vol. 31, no. 3, pp. 166–180, Jan. 2007, doi: 10.1108/03090590710739250.
[32] Y. Yehudi, M. Sharan, E. Tsang, B. Batut, and R. Mosbergen, 'Job vacancy: Community Researcher and Programme Coordinator - OLS is hiring!', Dec. 2022, doi: 10.5281/zenodo.7386879.
[33] MetaDocencia Team, 'Convocatoria laboral - Coordinación de Proyecto y Desarrollo de Cursos', Apr. 2022, doi: 10.5281/zenodo.7387015.
[34] MetaDocencia Team, 'Convocatoria laboral - Coordinación de Comunicación y Comunidades', Apr. 2022, doi: 10.5281/ZENODO.7387039.
[35] MetaDocencia Team, 'Convocatoria laboral - Coordinación de Infraestructura y Medición de Impacto', Apr. 2022, doi: 10.5281/zenodo.7387031.
[36] J. C. Williams and Dolkas, Jamie, 'Keeping Diversity Metrics While Controlling for Legal Risk: The Risks of Diversity Data are Changing', The Center for WorkLife Law, White Paper, 2022. [Online]. Available: https://biasinterrupters.org/wp-content/uploads/Keeping-Diversity-Metrics-While-Controlling-for-Legal-Risk.pdf
[37] U.S. Equal Employment Opportunity Commission, 'Job Patterns For Minorities And Women In Private Industry (EEO-1)', *US EEOC*, 2020. https://www.eeoc.gov/data/job-patterns-minorities-and-women-private-industry-eeo-1-0 (accessed Dec. 01, 2022).
[38] Bias Interrupters, 'Bias Interrupters', Sep. 13, 2016. https://biasinterrupters.org/ (accessed Dec. 01, 2022).
[39] N. Stevens, 'Open Demographics documentation — Open Demographics 1 documentation'. https://nikkistevens.com/open-demographics/index.html (accessed Dec. 01, 2022).
[40] T. Rajaram and M. Cantu, 'Metric: Inclusive Experience at Event', *CHAOSS*, Jul. 20, 2022. https://chaoss.community/kb/metric-inclusive-experience-at-event/ (accessed Dec. 01, 2022).
[41] TODO Group, 'TODO Group Code of Conduct'. https://todogroup.org/code-of-conduct/ (accessed Nov. 28, 2022).
[42] Otter Tech LLC, 'Code of Conduct Enforcement Training – Otter Tech'. https://otter.technology/code-of-conduct-training/ (accessed Dec. 01, 2022).

[43] The Carpentries, 'CODE OF CONDUCT — The Carpentries Handbook documentation'. https://docs.carpentries.org/topic_folders/policies/index_coc.html (accessed Dec. 01, 2022).
[44] The Carpentries, 'Code of Conduct Incident Response Procedure and Enforcement Guidelines — The Carpentries Handbook documentation'. https://docs.carpentries.org/topic_folders/policies/enforcement-guidelines.html (accessed Dec. 01, 2022).

4. Open-Source Science Development Plan

Virtual Cohort Training Program:

All training outputs will be made publicly available as they are created. Individual calls have both blank “call templates” and completed post-meeting call notes, which will be deposited online on GitHub and Zenodo, under a CC-BY 4.0 license. Slides from presentations are deposited on Zenodo. The calls are recorded, except for private group discussions in breakout rooms. All recordings are deposited on Youtube, again with a CC-BY license and full Spanish transcriptions to facilitate hearing, screen-reading and language accessibility. Where code and scripts are used to manage the program, all code content will be stored on GitHub under a permissive OSI Licence. Code content is accompanied by detailed documentation such that others can contribute to it. We also teach and encourage all project leads to license their project output under a CC-BY/permissive OSI license, and to document their work and plans openly and inclusively.

Virtual Cohort Program Impact Measurement:

We will measure the impact of virtual cohorts. The impact study will produce data which will be in the form of survey responses. We will share the survey protocols and methods openly on protocols.io or another relevant open service, including recruitment advertisements and survey questions. The study has been approved by an academic board for responsible research conduct.

Where ethically appropriate and safe we will share impact study response data anonymously or aggregated, as permitted by participants consent. Raw data will be stored securely on cloud-based servers (such as dedicated Dropbox for Business or GSuite) until processed (i.e. anonymised where necessary). Raw data will be retained for a minimum of five years unless ethics review or institutional repository requirements stipulate otherwise.

Once processed, data will be deposited on an open repository such as Zenodo or Dryad, alongside metadata “README” files that describe the shape of data and the meaning of specific columns.

Where possible, data will be stored in re-usable formats, i.e. CSV and JSON rather than proprietary formats, to facilitate re-use without the need for expensive or proprietary software.

We may use codes and scripts to analyze data, tidy the data, and produce visualizations. Where this occurs, the scripts and clear running documentation and dependency data will be stored on GitHub under a permissive OSI license, and versioned snapshots of the code will be deposited on Zenodo.

Results dissemination:

In addition to sharing the data, protocol, and scripts, we will ensure that all publications associated with our impact study and training program are posted as preprints to facilitate early comment and distribution, and finally in open-access journals.

This proposal:

This proposal will be deposited on Zenodo to get a DOI as well as shared via the Open Grants portal (https://www.ogrants.org/).

5. Equal Access Plan

Positive and inclusive virtual events: Senior staff in this project are highly experienced in inclusion related techniques (as recommended by [21], [22]). We will:

- **Anticipate accessibility needs:** applicants will be asked how their participation can be facilitated. Their needs will be assessed, processed according to accessibility profiles [16] and tackled based on individual requirements.
- **Language:** English predominance excludes those who do not speak, or have a limited understanding of English, from learning and contributing to science [24]. Spanish is spoken by 13.2% of the US population, being the most spoken language after English [25], it is the official language of Puerto Rico, and it is spoken in 20 other countries. Additionally, people taught in their mother language have higher retention [26] and better learning experiences. We will prioritize Spanish-speaking experts to help increase their visibility. But if not possible, we will provide high-quality live interpretation.
- **No-quibble fee waivers** for participants who otherwise would not be able to join.
- **Microgrants for cohort participation:** A budget of around $1,000 per cohort will be available for assisting participants with things like headsets, internet, childcare, or accessibility. Any surplus rolls over to future cohorts. We have experience making this process easy for recipients (e.g., transferring amounts ahead of time rather than reimbursing, minimal administrative layers and red tape). Microgrants enable higher uptake of often marginalized participants, and creates a sense of belonging to cohorts.
- **Hearing disability and language inclusion in online calls:** We will follow best practices for disability and language inclusion. Using real time AI-based captioning in Spanish during all calls will facilitate participation of, for example, those who are deaf or hard of hearing, or have ADHD.
- **Visual impairment and accessibility:** All training materials will use images with descriptive alt text for screen readers, and when using collaborative documents and the like, we will use visual-impairment tools based on recommendations such as the Web Content Accessibility Guidelines (WCAG) [27] and AbilityNet [28].
- **Ad-hoc access to training module calls:** Equal access to training also means flexible timing, as “free” time in itself is an unequally distributed resource [29]. All training calls will be recorded (excluding breakout rooms) and uploaded to YouTube, so participants can catch up if needed. Automatic transcriptions will be error-corrected before upload [22] and online ad-hoc drop-in sessions will be offered for people to interact with their cohort.
- **Diverse leadership, staff, instructor, and mentor team:** People are more likely to see themselves as members of a community if there are others like them [30], [31]. Our instructors and call hosts come from a variety of backgrounds and career paths, both

traditional and non-traditional.

- **Building a sense of community via shared ownership:** Once participants have participated in one cohort, those assessed as ready by their mentors will be invited to become mentors. Mentors are paid an honorarium, to ensure that those who are more time-poor are still able to prioritize mentoring if they wish to.

Training: Mentors and staff will participate in Ally Skills and Bias + Inclusion training sessions, learning to recognise and use their power to become allies of marginalized people. Also, online call facilitators are trained in inclusive techniques to work with people with diverse modes of communication, language skills, and ability/disability, covering visual & hearing impairments and general best practices.

Positive and inclusive working environment: The conduct guidelines apply equally to staff, as do our virtual inclusion plans. We also plan to follow good inclusion practices for recruitment and staffing [32]–[35].

Metrics: Reducing and avoiding overt and unconscious discriminatory bias is core in our proposal [36], [37]. Using metrics will allow us to observe inadvertent patterns and act to rectify them (implementing “bias interrupters”[38]). Some core measures will include gender, sexuality, race, and disability, drawn from open source demographic standards [39], inclusion standards [40], as well as indirect metrics such as a participant’s and their parental income, and education. All measures will comply with mandatory federal reporting standards for employees, and be optional but encouraged for participants and mentors. The indirect measures will be designed specifically to understand a participant’s family background, with a goal to take affirmative action (legal and appropriate) and thus increase not only diversity but active participation from participants with disadvantaged-socioeconomic backgrounds.

Conduct principles: Our code of conduct (CoC) and participation guidelines will be initially based on the Todo group [41]. Once we have onboarded mentors, experts, trainers, and staff, we will hold workshops to develop community-derived practices, using existing recommendations from organizations like Otter Tech, code of conduct violation response experts [42], and the Carpentries [43]. Multiple reporting options will be provided, in case the person being reported is a CoC reporting contact.
The responses will be proportionate and will consider the following: likelihood of re-offense; danger to community (if any) due to either inaction or overharsh action; effects of investigation on the reporter and the reportee; speed in responding to CoC events [44]. Equally, we commit to not employing such haste that people are inadvertently hurt during the process.