

**NOTE TO READERS:** This transcript has been edited for readability and may deviate slightly from the actual recorded conversation. If you have any questions, please contact the OED working group: [BIO-OED@nsf.gov](mailto:BIO-OED@nsf.gov)

I want to welcome everybody to the webinar today about the Center for Advancement and Synthesis of Open Environmental Data and Sciences. There are a few housekeeping items first, and you should see a notice on your screen that this meeting is being recorded. You can hit the continue button to continue in the webinar or leave if you don't want to participate while it's being recorded. We will record the webinar so that material is available later and will post it on the website shortly. Our speaker today is Matt Kane, and Matt will do the main part of the presentation, which should last for about 15 to 20 minutes, and we have left as much time as we can to entertain your questions about the synthesis center moving forward. All webinar participants are in "listen only" mode today, and we have disabled the chat. We have enabled the question-and-answer function, and you should see the small Q&A button at the bottom of your screen. We would ask you to submit any questions you have during the webinar through the Q&A function. Because your name would otherwise be visible to the other participants, make sure if you don't want your name shown that you set the question to be asked anonymously, which is one of the functions available when you type in the question. I think that covers the major housekeeping things. We will turn it over to Matt to let him continue with the discussion.

Thanks, Steve. Good afternoon, everybody. Thank you for joining us for this webinar. We are very excited about the solicitation for the Center for Advancement and Synthesis of Open Environmental Data and Sciences. We have been working on this now for a couple of years. This is a very exciting competition, and a very exciting time in environmental science. This solicitation is a collaboration between the Divisions of Biological Infrastructure and Environmental Biology, in the Biological Sciences Directorate, and the Office of Advanced Cyberinfrastructure, in the Computer and Information Science and Engineering Directorate. The working group for this solicitation and activity includes myself, Matt Kane, in the Division of Environmental Biology; Steve Ellis in the Division of Biological Infrastructure, who introduced me; Tevfik Kosar in the Office of Advanced Cyberinfrastructure; Reed Beaman, Division of Biological Infrastructure; and Liz Blood in the Division of Environmental Biology. We have emails provided here and will be sharing them at the end of the presentation. Feel free to contact us with any questions or if we can be of any help.

Open environmental data is burgeoning in the Biological Sciences Directorate, where we especially think of it in terms of the National Ecological Observatory Network, which came online almost 2 years ago now. It has 180 some different open data products, available to anyone that wants to use it. There are many other sources of open environmental data, such as the Long-Term Ecological Research Network (LTER), which is also funded by the Biological Directorate, the Critical Zone Observatories program, and the National Center for Atmospheric Research, another NSF center, as well as other sources such as the NASA satellite program, most recently involving the GEDI system, and others. There are others supported by NSF, such as iDigBio and the Global Biodiversity Initiative, so we have many sources of open environmental data available. I have really only mentioned some of what is out there. Certainly in the Biology Directorate, these are the ones on our mind, and we have felt that the user community could really use a center like this to help to bring people together and move the field forward.

The center we are thinking of will have four pillars upon which it will stand. I will go into each of these in a minute on the next slide. The first is it is an incubator for team data science. It is also going to involve creative and innovative cyber infrastructure. Additionally, the data is open to everybody, which has a wonderful democratizing impact. It will be very inclusive, with community engagement focused on broadening participation in environmental science. The center will also have a major data science training component.

Let's unpack these a little bit. The center will serve as an incubator for team-based, open data-driven research in environmental biology through support and coordination of groups developing new questions, approaches, analyses and syntheses. The center will engage a diverse and inclusive scientific community, broadening participation of underrepresented groups in science, and it will democratize data science, and support a diversified STEM workforce. It is going to provide the needed software, data and compute for research through development and support of creative and innovative cyber infrastructure, tools, services and applications that facilitate accessing, manipulating, analyzing and visualizing data from multiple sources to address novel research questions. And then finally, the center will feature immersive training and research experiences for diverse undergraduates, postdoctoral researchers, scientists and educators at all career stages, and provide access points for faculty at both research- and teaching-intensive institutions. Those are the four pillars upon which we envision this center thriving.

In terms of staffing, the key positions involve first and foremost, the Center Director. They will be responsible for scientific leadership, and are expected to have management qualifications which are critical to the success of the project. The Director is responsible for management, staffing, and ensuring an inclusive and equitable climate that will welcome diverse students and users. The PI of a proposal should be the Center Director or Interim Director. Note that Co-PIs are not permitted.

Another important role is the Project Manager or Associate Director position. This individual will oversee most of the day-to-day operations of the Center, including procurement and management of resources, scheduling, assessing risks and opportunities, and ensuring that milestones and outcomes are being met, and budgets are balanced. There are other key positions that also may be involved, depending on the PIs and teams that submit proposals. For example, the Center could require full-time staff positions in areas such as education and training, cyber infrastructure and data management, and others to ensure that the center is successful in carrying out a visionary and thriving program.

In terms of Senior Personnel, this solicitation has three different stages, each with its own staffing instructions. The first stage is a Letter of Intent. You can't go to either of the two subsequent stages without submitting a Letter of Intent. NSF guidelines permit a maximum of four Other Senior Personnel in the Letter of Intent, in addition to the PI. In the Letter of Intent you should list all participating organizations. The Letter of Intent as I mentioned, is a requirement for submitting a Preliminary Proposal, which is the next stage. The solicitation indicates that a two-page biographical sketch should be included for the PI and up to five members of the leadership team. These are aspects that are hardwired into the NSF submission system, which is why we are making sure that you understand them. Preliminary proposals must also identify all participating organizations. I also want to mention about the Letter of Intent - originally the solicitation indicated that the synopsis and the letter was limited to 500 characters. That is incorrect, and we have removed it from the solicitation. The form will permit up to 2500 characters for this synopsis in the Letter of Intent. Finally, the last step is submission of Full Proposals, which are only to be submitted if invited after peer review of a Preliminary Proposal (so only invited PIs & teams can submit Full Proposals). The list of participants should include the PI and all Other Senior Personnel. There is no maximum on the number of people that you indicate in the Full Proposals. The two page biographical sketches should be included for the PI and all Other Senior Personnel.

The timeline for each of these is as follows. The Letters of Intent, which are required, are due April 1. The Preliminary Proposals are due April 29. These are deadlines and not target dates. We hope to have invite/not invite decisions finalized by June 15. Then, Full Proposals are going to be due on September 15, five months after the preliminary due date and three months after the decisions on Preliminary Proposals. Then, following the Full Proposal panel, a few proposal teams will be selected for site visits. These will be scheduled in October and November. Finally, we hope to have an award selected by the end of this year, by the end of 2021. In terms of funding, this will be a five year award with the budget for year one for the

center at \$2 million. Year two is ramping up to \$3 million. It will ramp up to \$5 million/year for the next three years. The funding will be via a cooperative agreement and the center will be eligible to apply for renewal for a second five-year award.

At this point we are ready to entertain any questions.

Please use the Q&A button at the bottom of your screen and remember to set it to ask anonymously. We have a first question coming in. I will read them out, Matt and other members of the working group can chime in as you want to answer it.

**To what extent is the center expected to pursue its own synthesis questions akin to NSF or is a process for the broader community to include their questions?**

I think the primary function of the Center is to engage and serve the community of scientists, working with open environmental data. One can certainly do their own research as part of it, but the function of the Center isn't primarily to do one's own research. The Center Director and their team are going to be fully occupied with engaging and training the community of scientists in this area.

**For the next one, is NSF cooperating with NASA on this center and can an existing center apply for the solicitation?**

NSF is not cooperating with NASA on the center, but we certainly interact with NASA a lot in various ways. In terms of an existing center, we don't have a general rule against concurrent award for any individual award at least, but you would have to argue feasibility and ability to address the center goals in your proposal. If you're asking whether a NASA center be able to apply, I think the best answer is that Federally-Funded Research and Development Centers (FFRDCs), which NASA and other agencies fund, are not eligible to apply directly for NSF support. They can serve as sub-awardees on an NSF proposal.

**Next question, will the renewal have the same graduated financial structure?**

At this point, that is more than five years away and we don't have an answer that is definitive. But one can imagine that funding would likely work in reverse. That is your six, seven and eight years would be at the \$5 million level, and it would begin to ramp down in years nine, and then ramp down again in year 10. That is typically what we have done.

**Does the PI need to be a data scientist for the application to be favorably reviewed?**

No, the PI does not need to be a data scientist to be favorably reviewed. You need to write a convincing proposal. We don't review the Letters of Intent of course but the Preliminary Proposals need to be convincing to the reviewers and everyone else. That is the major consideration, is that you are able to demonstrate that you are qualified as the PI and that your team is best qualified.

**For the LOI synopsis does the character count include spaces?**

Yes, it does include spaces.

**The next question: is it either advantageous or disadvantageous from the NSF perspective if a proposed center consists of universities geographically proximate or distal to the primary institution?**

I think that is really up to the PI and their team to convince people in writing their Preliminary Proposal as to what is the optimal logistical configuration for the Center, and that the way that they want to do it is optimal, whether that involves geographically proximate or distal organizations. We have no predisposed idea of what the center should be in that regard. I think it is up to each team to convince reviewers that your solution is the best.

**Is there a recommendation for the long-term funding model for this center? I assume it is supposed to last into perpetuity.**

I think that assumption is not the NSF assumption. Typically NSF supports centers, at least in the Biology Directorate, for a period of ten years, if they are successfully renewed. If the center continues beyond that, it has to find other sources of support. I don't think that any of us can really predict where open environmental data science is going to be 10 or 11 years from now, and whether a center that is created will serve the same needs that the community has today.

**The next question: would a geoscientist PI without a main background in biological sciences be considered competitive for this call? In other words what should the balance of scientific leadership, skills and management, and open science be?**

Again, I think this is kind of an answer that is determined by the review process. It is up to each PI and team to write a persuasive proposal that convinces the readers that they have the experience and knowledge, and skills, to co-manage the center more than any other team. There's no reason why that could not be a geoscientist or computer scientist, or a biologist. I will say that the primary thinking in the biological sciences directorate is that our big investment in open environmental data is the National Ecological Observatory Network. We want a center that will certainly address the needs of the NEON community. That is more or less a requirement of the center, but there are many other sources of open environmental data, and NSF has other investments in this area. However, the kind of data that NEON is producing is of primary interest to us. Advancing environmental biology as a science by using open environmental data is also of primary interest here.

**Are there any network data sources, or infrastructures, that are essential to include in the center?**

I think we just said that we certainly consider NEON and the data that it is producing, and the science it is going to produce, as being essential to include in the center. But we don't at all envision that the center will be limited to NEON. That is why in the introduction we mentioned all of these other sources of open environmental data.

We envision the center to be able to support access and analysis of diverse and specific data sets, including those from environmental and biological data sources. In order to achieve this, the center needs to enable user access of different data, networking, and computing sources. We would like the center to use as much existing resources as necessary in achieving the Center's goals. The Center may be a need to develop some new CI components in order to utilize the existing data and assorted resources but should not "reinvent the wheel."

**Related open data and environmental decisions require knowledge of co-occurring social economic features. With the NSF approve of adding social economic data to this mix?**

It is up to the proposers to convince the reviewers that the right mix of scientific focus is what you are proposing. I can imagine important environmental biology science questions involving social and economic features. In fact, for many years now the biological sciences directorate has collaborated with other directorates in this area.

There are certainly opportunities. Again, depending on the questions, working groups that are part of the Center to address those questions could look at the human dimensions as well as the ecological or physical sciences dimensions. We certainly provide funding through NSF through several programs looking at those aspects.

**The next question is given the COVID experience, can you envision this as a distributive virtual center rather than a bricks and mortar institute?**

We envision that the Center will have both in person and virtual components. It is up to each proposer to put in their proposal what they think is the optimal mix to meet the center's goals and objectives as we have articulated them in the solicitation.

This also relates particularly to the concept of democratizing science. It is envisioned that the center is going to provide opportunities for not only underserved communities, at small community colleges, but opportunities for lifelong learning for folks. Those individuals that are not going to have the money and resources to be able to travel to the Center need to be able to participate, so to some degree there will be a need for a virtual component if you are going to be successful in reaching the overall goals of the Center.

**Next question: To what degree should the center proposal be motivated by specific research questions? How much effort and space should be devoted to elaborating upon these questions, and specific approaches for addressing them?**

We envision this Center as really addressing many questions that are coming from the community in the form of the science incubation teams that the center will support. Undoubtedly, the Center Director and their team will need to provide some scientific leadership in this area as well. But the Center is not meant to support the research of the PI and their team primarily. It is actually meant to support the community and to include the four different areas that we discussed at the beginning of the webinar.

**The next question is, to what extent would you like to see the center engage beyond the scientific community to the end users, and outcome predictions generated by syntheses?**

This fits into the area of broader impacts of the NSF funded research. Every proposal we evaluate is evaluated on both intellectual merit and broader impacts, including the Center. As Matt indicated it is really up to the proposers as to what your focus would be, what emphasis you would initially have, or how you would accommodate such a broader impacts as part of the activities, whether that is training or synthesis that your center will focus on.

**The next one goes back to the question about NASA, and the questioner is saying they did not mean the NASA center but just an existing center. So does the proposal need to establish the center from scratch, or could it be led by group of people already organized as a center?**

Certainly it is an open solicitation with the usual qualifications that all NSF proposals have. A team from an existing center at an institution can certainly submit a proposal.

**The next question is, will data science training be available to anyone that wants it?**

That is certainly our goal, that is the democratization of environmental science through open environmental data. That is what we consider one of the fundamental missions of this center.

**Is the lead institution expected to have all of the experience that are expected for the Center ? For example, the science expertise, training and education expertise, and cyber infrastructure expertise?**

I would answer that the Center, in whatever form it is in, whether it be a single institution or multi-institutional endeavor, needs to provide all of the necessary computational and infrastructure and knowledge, data, environmental science knowledge, team science, the training, with a critical aspect being diversity and equity. All of these need to be present as part of the Center, be the center a single institution or multi-institution. If it is a multi-institutional endeavor, it is conceivable that different expertise would be represented by different institutions.

**What are the ideal outcomes of such a center from the NSF point of view? What problems in the scientific community does NSF feel it is addressing with this solicitation?**

All of the types of questions and problems that are addressed in environmental biology and broader environmental science are meant to be served by the Center. Those problems that can be best addressed with available, open environmental data, and with the tools that teams that will be involved and will be Center-developed are those that the Center will provide solutions to. We don't have specific questions that the Center is charged with addressing or answering. The Center's outcomes will be very broad.

**What are the expectations during the proposal writing stages that the infrastructures mentioned in this solicitation are they expected to be neutral?**

This question is unclear; we will move on and hopefully that person can expand on what they mean.

**Can existing historical NSF centers or locations apply for this solicitation ? For example, NCEAS?**

Certainly existing centers are able to submit proposals to this competition as well.

**The Center is expected to deploy and manage data distribution systems and act as a facilitator between existing centers and the need to researchers?**

We expect the Center to facilitate the findability, accessibility, interoperability and reusability of open biological and environmental data from distributed and heterogenous sources for researchers to address novel research questions. In achieving this, the Center will provide the researchers with the needed software, data, and computing cyberinfrastructure (CI), tools, and services. Whenever possible, the Center will leverage existing CI components in lieu of acquiring or developing new ones.

**There has been some discussion on our campus about the four pillars. None is more important than the others, correct?**

I think that is accurate. We are envisioning a center in which all of these four pillars are seamlessly interwoven with one another in the development of this center. We are interested in your visions of the mechanisms through which that is going to happen.

**Is BIO cooperating with GEO in organizing this center? What is the relative importance of focusing on BIO funded data, for example LTER sources, versus the ones funded through GEO, such as NCAR?**

The Biology Directorate and Geosciences Directorate collaborate on a lot of things. The Geosciences Directorate is not currently a signatory on this solicitation. The emphasis here is on environmental biology, so biology is front and center in the environmental science for the center. Certainly we are in close communication with our colleagues in the Geosciences Directorate. It is possible that either as this center evolves or as the competition evolves, that the relationship could change. But for now it is just the Biology Directorate collaborating with the Office of Advanced Cyberinfrastructure.

The working groups' focus will determine to a large degree the data resources that are needed. NSF does not have any preferences or biases associated with the data to be used or where it comes from. The Center should use the open environmental data that is critical to be able to address the synthesis or questions that are being asked.

**What are some models of how the center will interact with entities, for example, NEON, mentioned in the first few slides?**

I think there are no models on how the center would interact with these entities, which is why we issued the solicitation. While the Center builds on the Biology Directorate's history of synthesis centers, this is a new and different center that is meeting the needs for the first time.

**The next question comes from a PhD student that is interested in how we expect to have students engaged. Will these opportunities be for existing NSF fellow students, from your own institution or region, or open to all students?**

I think it is much too early to understand exactly how different students at different levels will be engaged, but we do expect they will be engaged. It is up to the proposers to help craft a vision for the center that describes how the user community, including graduate students, will be involved, and how they will support training for graduate students, for post-docs, and for many others.

**What is the rationale behind limiting the center to one PI, and no Co-PIs?**

Clearly the center needs a strong leader, a visionary leader. In our experience in running synthesis centers, this is the model that we have found is the most successful. The Biology Directorate at NSF, as indicated in the solicitation, has a history of supporting centers for analysis and synthesis dating back over 20 years now. This is the model we have found works best.

It also goes toward accountability. It doesn't lower the expectations or the importance of the other senior personnel associated with this center.

**What do you anticipate the approximate budget balance between participant support for working groups versus technical and administrative staff, for example cyber infrastructure, versus center supported postdocs and students?**

This is a vision we would believe would be provided by the Center Director and their management leadership team. Different people will have different solutions to this. It is up to each team to convince people in writing their proposals why their solution is optimal.

**How is this center expected to interface (or not) with groups such as OOI who already have and continue to develop open and practical access to their data.**

I think this is going to depend on the Center PI, Director and leadership team as to the mechanisms that the center will use to interface with NEON, OOI and other groups and other teams that might be engaging in open environmental data analysis and synthesis. That is going to be the role of the Center Director and their leadership team.

**Where will data and modeling outputs generated from the award be shared at the conclusion of the project?**

It is really up to the PIs to decide where they want to store the project data and what they want to share. We expect the data to be available to share to the community and we expect PIs to come up with strategy to store and preserve the data, and to share it with the stakeholders.

**Next question, does NSF frown on an individual involved in a previously funded center being involved in this center?**

We have no preconceived notions of the specific center experience that the Director for this center brings to bear. Simply put, no, we do not frown on that.

**Is there an ideal relationship envisioned between the center, NEON and other NSF funded observatories or depositories? Would they be consultants and contractors ideally?**

I would view the relationships as establishing partnerships and establishing collaborative relationships, rather than as consultants or contractors. Although it is conceivable that as organizations involved in other NSF funded activities could put in proposals, they could also certainly serve as consultants or contractors. Fundamentally, we see this as a collaborative relationship.

**The next question is what kind of partners should be explicitly identified? If one wants to partner with those networks explicitly named in the RFP, such as NEON, should they be collaborators or have subcontracts?**

It is up to each individual PI and team, and will depend on the nature of the collaborative relationship. If their relationship with more than one facility or network is going to involve financial aspects, or if they are going to be involved in subcontracts or sub awards, then they will need to be listed in the letter of intent, and in the pre-proposal and full proposal. How they are identified will depend on the nature of the relationship.

This is something we expect the proposers to really lay out in their proposal. If there are specifics about collaboration, then yes, it should be noted. But if they are just generally accessing resources that are available to the public, they don't need to be listed as a collaborator.

**Will the slides be made available for download?**

Yes, we intend to make the presentation available publicly.

**Are there expectations for how funds should be allocated in terms of percentage to administrative, research training and cyber infrastructure support?**

No, there are not specific expectations.

**Is it encouraged or advantageous for lead institutions to work with partner organizations for specific components of the project objectives? Such as collaborating with the STEM education center for data science training, etc.**

It's up to each proposer, each PI and team to decide what partnership organizations should be involved to create the optimal center to meet the objectives and goals as described in the solicitation. We have no preconceived notion of the combination here.

**NSF has invested considerable funds to organizations and institutions to develop tools and cyber infrastructure to harness and disseminate data. How does NSF envision the center will best leverage existing and past efforts, not reinventing the wheel, but to move forward that is complementary but not competitive?**

We expect the PI to deliver the existing components, and to identify the gaps in between. We assume they will need to leverage existing CI, but the existing CI may not be sufficient to achieve what they want to do. We expect them to identify the gaps in terms of the CI delivery.

**The next one is should we expect to fund NEON workers through this funding opportunity ?**

NEON employees will not be supported through this funding opportunity for the operation and maintenance of NEON facilities. [Clarification: If NEON employees would be making contributions to the Center that fall outside the scope of those required for NEON operations and maintenance then they can be funded for their time and effort through the Center.]

**Given the scope of the environmental databases outlined in the presentation, would a nationally focused center be prioritized over globally focused one?**

To some degree the question has two parts. One has to do with NSF policy related to supporting international institutions. The NSF funds are intended primarily to support U.S. institutions. An important focus of the Center is on data generated by NEON, which is a US facility. You can, through partnerships, if there is a very unique resource or capability that is not available in the U.S., justify some potential international partnerships. There are some restrictions in terms of NSF dollars going to international institutions or international individuals. That being said, there are a lot of critical questions that require a global focus, to be able to answer. Therefore it would be reasonable to access any environmental data from any source around the world, including satellites, including data from other countries, as part of the center itself. There may be interesting resources that would be available to the larger community.

**The next question: given the competition for the NEON operations program this year, how should the environmental synthesis center propose to coordinate with NEON, and could the direction change significantly under a new operator? When will we know who will operate NEON?**

Questions about NEON and the competition should be directed to the NEON program director at NSF, Roland Roberts, and he can provide the details about competition on NEON. Whatever organization is managing NEON operations should not impact the goals and directives of the Center.

**Should the center be expected to purchase its own computer hardware or can it rely exclusively on existing cyber resources?**

NSF has already made significant investment in computing and harvesting data. This includes other leadership, class and facilities and many other computer resources. We expect the center to leverage these resources as much as possible. If the center decides it needs very specialized hardware not already provided by the existing NSF resources, then we will be open to provide support for such hardware in specialized cases.

**Can you clarify what you mean by democratizing science when talking about the center? Is referred to education, governments or research activities?**

What we mean is that open environmental data is available to anybody. It has the power, therefore, to be accessed by people at large research universities, or teaching focused universities, or high schools, or other institutions and research centers. Anybody can access it. That is what we mean by democratizing science. Because the data is available to anybody, anybody who has the knowledge and abilities can use this data to ask questions about environmental, biology, and environmental science. When you say does it refer to education, governments or research activities, it refers to all of those, all of the activities of the Center are really focused on creating an inclusive and equitable scientific community in this area.

One of the challenges of dealing with data from the observatories like NEON is understanding how to access it, and process it, so the training is really important. But not every institution has the faculty or capabilities to conduct the training needed to be able to avail yourselves of this data. One of the things that we envision the Center doing is to provide training to allow people to be able to access it, and use it in ways that they might not have otherwise because they would not have had access to the training to be able to use it. That is another aspect of democratization, giving people the skills and capabilities as well as equal access to the data.

**How much of the proposal should be devoted to describing the scientific research problem versus focusing mostly on training and science, support activities; would proposal be more like the NRT proposal or the BII proposal as examples?**

I think it will involve elements of both. NRT is the National Research Traineeship program, and this proposal certainly will involve major training aspects. BII institutes are completely focused on the research theme. I think with this Center here, the community of researchers should be enabled to bring their research questions to the center in an open way. That openness is the distinction.

We think of this Center as an enabling center for both the scientific questions and the scientific community. Clearly that is going to involve a great deal of knowledge and expertise in leading the Center. But it is not a center for someone or a team of people simply to do their own research and their own work. One can imagine that their own research and work will be part of the Center as well, but the focus is on enabling science, enabling the community.

Remember that you can always send us a question via email. I have pasted our alias in the chat, BIO-OED@NSF.GOV. With that, we are ready to conclude the webinar. Thank you for your attendance and excellent questions. We look forward to seeing your LOI's, pre-proposals and full proposals. Thank you all for tuning in.