Response to NSF's Request for Information on the CHIPS and Science Act Section 10343. Research Ethics

On 08/27/2024, the National Science Foundation published a Request for Information (RFI) to seek input to inform the development of the agency's response to Section 10343. Research Ethics in the CHIPS and Science Act of 2022.

Below are the questions posed by NSF and the responses submitted by David H. Guston, Associate Vice Provost for Discovery, Engagement and Outcomes, and Arthur Daemmrich, Director of the Consortium for Science, Policy and Outcomes (CSPO).

Ethical, Social, Safety, and Security Considerations

Question 1: **Describe ethical, social, safety, and/or security risks from current or emerging research activities that you believe might be of concern to the community, profession, or organization with which you are connected.**

Summary: "Risk" has a formal meaning that is too narrow to include the important scope of ethical, social, safety, and/or security" (ES3) concerns. NSF should articulate its own relevant definition of risk or find another term or concept.

I am very much in sympathy with what I understand is the perspective of the relevant legislation (CHIPS and Science Act), that is, all areas of research when closely examined have "ethical, social, safety, and/or security" (ES3) risks. Many of these risks are cross-cutting and technology agnostic, e.g., "will their upsides and downsides be equitably distributed across (the relevant portions of) society?" Others may be highly specific to the project or the research domain, e.g., "is it appropriate to build this large-scale research facility (e.g., telescope) in this place (atop a mountain sacred to an Indigenous culture)?"

In my 11 years directing the NSF-funded Center for Nanotechnology in Society at Arizona State University (CNS-ASU; #0531194 and 0937591 plus various supplements, 2005-2016), I learned that many but not all cross-cutting risks may be identified *a priori*, but also that many and perhaps most of the specific risks require close, interdisciplinary inquiry to discover. We also learned that the risk framework or risk paradigm itself – in which risk is formally conceived as the product of a hazard and its probability of occurring, or less formally conceived as a balancing of benefits and disbenefits – is inadequate in almost if not all research-related circumstances. For example, the risk paradigm does not deal with distributive issues, it has a hard time with inter-generational issues, it cannot by definition manage issues that are not quantifiable and thus it cannot manage emerging technologies for which there are no data about first or second-order community or societal impacts. Therefore, prior categories in common parlance, e.g., "unanticipated risks," need to be not only revisited but perhaps discarded, both because anticipation itself is an effort to be made through systematic inquiry (see below), and also because the category of risk is not capacious enough to handle many necessary and interesting things.

If NSF continued with risk language, it would need to articulate a definition of risk that is not dependent on the formal, quantitative definition and is not bounded by a cramped vision of "balancing."

***Question 2:* Which products, technologies, and/or other outcomes from research do you think could cause significant harm to the public in the foreseeable future?**

Summary: AI is clearly the technology that requires the most attention, but there are critically important issues across emerging technologies – and many of these issues are common to most or all emerging technologies. It is therefore hard to prioritize one technology area over another, and NSF's approach should not prioritize AI to the detriment of other approaches.

From my CNS-ASU experience, my long-time work with the Consortium for Science, Policy and Outcomes (CSPO), and my leadership role in establishing ASU's School for the Future of Innovation in Society and operating ASU's Julie Ann Wrigley Global Futures Laboratory, I find the framing of this question somewhat challenging. First, as above, it is not just the "products, technologies, and/or other outcomes" that could cause harm. There are some research *questions* that some communities, rightly or wrongly, view as harmful (e.g., human origins research applied to Native American populations), and there are clearly some research *practices* that cause harm, whether intentional or not. Second, "the public" is not a unitary entity. Research and its related products, technologies, etc., most often create both losers and winners, and it is appropriate to ask, throughout the research process, "who are the winners and losers if it is done this way?" and to attempt to adjust those anticipated outcomes in ways that meet the explicit and implicit goals of the funding program, the reflected values of the investigators, any relevant communities interacting with the researchers, and a broader and empirically defined set of public values (e.g., the organic act of the sponsor, relevant executive orders, and other such public and authorized expressions of intent, mission, and perspective).

If by asking questions 1 & 2 here, NSF is looking for ways to prioritize some areas of research for specific attention, I would argue that it is hard to know without close examination, and even people with close experience in one field may not have close enough experience with another field to make rigorous, comparative judgments. Moreover, the types of concerns represented even in the ES3 categorization are incommensurable, e.g., making trade-offs between anticipated ethics issues and anticipated security issues very challenging. Moreover, harm is not a fixed category, as harm to an individual could benefit a community, and vice-versa. (This is to say again that technologies have winners and losers). I conclude from these points that a broad-based capacity to ask and explore such questions – what CNS-ASU called "anticipatory governance – is necessary to establish in as broad and deep a way as possible.

That being said, the shear scope of change being wrought by AI – from impacts on markets and energy generation and mental health and well-being, to opportunities for useful discovery and analysis – seemingly demands significant attention. Such attention should not, however come at the expense of attention to areas like synthetic biology, geoengineering and climate tech, etc.

Question 3: **Describe one or more approaches for identifying ethical, social, safety, and/or security risks from research activities and balancing such risks against potential benefits.**

Summary: With large-scale NSF support, the Center for Nanotechnology in Society at Arizona State University designed and piloted approaches of real-time technology assessment and anticipatory governance that were meant to address precisely these issues across emerging technologies.

CNS-ASU built a set of methods (real-time technology assessment – Guston and Sarewitz 2002) and a vision for their application (anticipatory governance – Barben et al. 2008; Guston 2014) that, through a combination of foresight, public engagement, and interdisciplinary integration, help identify concerns about emerging technologies and begin to plot a path toward addressing those concerns. While initially applied to nano-scale science and technology – CNS was part of the Nano-scale Science and Engineering Center (NSEC) program at NSF – anticipatory governance is designed to apply, and has been applied, to other emerging technologies like geoengineering (e.g., Gupta et al. 2020), synthetic biology (e.g., Gorman 2012), and AI/robotics (Boyd and Wilson 2021), among others. Making use of scenario-based foresight activities, it is specifically designed for areas without good empirical information to plug into the risk paradigm. It is specifically adapted for "upstream" public engagement that helps publics understand – and contribute to – the goals of research and its application. And it is specifically adapted for "mid-stream modulation" that sees the laboratory as a place of societally relevant decision-making and offers opportunities to make such decision-making more deliberate and informed (that is to say, reflexive). In addition to initiating such activities upstream in the process of discovery and innovation, the real-time technology assessment and anticipatory governance approaches differ from traditional approaches like informed consent because the latter framework is limited to our ethical obligations to individual research subjects and not as well to governance issues addressing social issues.

The vision of anticipatory governance has already taken hold in many places. The OECD (2024) recently issued a report endorsing the anticipatory governance of emerging technologies. The "AREA" framework – anticipation, reflection, engagement, and action – deployed by the Engineering and Physical Sciences Research Council in the UK is largely derived from anticipatory governance. British Standards Institute constituted a publicly available specification (PAS 440) supporting responsible innovation for private sector innovators that draws heavily on engagement and anticipation. An international academic/professional community has grown up around anticipation, including UNESCO chairs in anticipation and anticipatory governance and a biennial conference. New areas of applied research, e.g., "anticipatory life cycle assessment" (Wender et al. 2014), have opened up.

Question 4: **Describe one or more strategies for encouraging research teams to incorporate ethical, social, safety, and/or security considerations into the design of their research**

approach. Also, how might the strategy vary depending on research type (for example, basic vs. applied) or setting (for example, academia or industry)?

Summary: Strategies include stand-alone center-level funding for ES3 work, integrating ES3 work into other large-scale science and technology research, supporting individual investigators, supporting networks, and supporting graduate training. International dimensions of such work is crucial as many important ES3 issues have global dimensions.

In addition to directing CNS-ASU, I was principal investigator on a workshop award (#1445903) on "Research Agendas for Societal Aspects of Synthetic Biology," with co-PIs Richard Murray (Caltech) and Jennifer Brian (ASU). Among the conclusions of the workshop was the strong message from the community assembled that synthetic biology – just one among several emerging technologies that NSF invests in – required large-scale, stand-alone funding for societal research (akin to the centers for nanotechnology in society at ASU and UC-Santa Barbara) and integrated funding in other large-scale synthetic biology investments (e.g., as in the Synthetic Biology Engineering Research Center), as well as individual investigator initiated societal work that might or might not be integrated with the biology.

Encouraging the incorporation of ES3 activities in new, large-scale, stand-alone centers is simply a matter of competing the centers and thinking about how, instead, to encourage such centers to reach out to their science and engineering colleagues. NSF could:

1. Create one or two such centers for each area or emerging technology it wanted investigated (e.g., each of the tech areas defined by the CHIPS and Science Act could have its own ethics- and society-driven center). Given the work that CNS-ASU performed both separately from but also occasionally in collaboration with CNS-UCSB, there is certainly enough work for two centers per tech area.
2. Define as requirements or desiderata some particular, collaborative modes of action within such centers, e.g., the co-funding of graduate students and/or post-docs. In CNS-ASU, we provided 1/3 of tuition and stipend to doctoral students in science and engineering in exchange for the time they spent with the center. Such students served as excellent conduits of exchange and performed their own research, including the incorporation of societal perspectives into their doctoral dissertations and opening up new areas of inquiry (e.g., anticipatory life cycle assessment, public engagement, etc.).
3. Endorse particular methods for one, some, or all such centers, along the lines of OECD's endorsement of the anticipatory governance of emerging technologies. For example, one design approach would be to create a network among three centers – one dedicated to each of the capacities of futures, public engagement, and interdisciplinary integration. Or each tech-specific center might develop all three capacities, as CNS-ASU did.

Encouraging the incorporation of ES3 activities into large-scale science and engineering activities is quite different, and lessons learned from a variety of activities, including in Europe as well as the U.S., is that such collaboration should start as early as possible in the development of STEM research agenda or activities. Given the experience of the NSF SynBERC (where I was an external reviewer), grafting even highly esteemed social science onto an existing research endeavor is not a recipe for success. While my colleagues and I have seen solid collaborations develop from early-stage partnerships between scholars of anticipatory governance and STEM researchers, we have also seen frustrations on the part of the societal researchers in not being able to establish their own vision of engaged research or to inflect the trajectory of the overall research in which they are collaborators because their numbers are too few and their structural position in the collaboration is too dependent on the STEM work.

Encouraging such research at the level of the individual investigator is the most straight-forward of these three, as NSF has had intellectually robust if very modestly funded programs in areas cognate to ES3 for a significant time. I believe that the CHIPS Act mandate should encourage NSF to more closely consider the expansion of such programs to address the expansion of ES3 challenges – many of which have been driven by NSF's own success.

Beyond stand-alone centers, required integration with technically-focused centers, and individual investigator approaches, NSF should also consider the need for focused community building through networks and through graduate training. CNS-ASU, while centered at ASU, created a network of researchers that, across 11 years, included researchers at half a dozen other domestic institutions of higher education in very substantial ways and included scores of other IHEs, for-profit and not-for-profit institutions as partners in a variety of activities. CNS-ASU supported numerous students and contributed to scores of research theses at the undergraduate, master's and doctoral levels. But to expand its impact, it created a cohort-oriented winter school that – with support from NNIN beyond CNS-ASU's lifetime – will have trained more than 170 early career researchers by January 2025.

Twenty to 30% of these early researchers were also from outside the United States, and over its 11 year lifetime, CNS-ASU hosted more than 100 international scholars (beyond winter school) for short- and long-term stays – in part because CNS-ASU conceived of itself as a "user facity" for training in the techniques of real-time technology assessment and anticipatory governance. The ES3 issues that CHIPS is concerned with are global issues, and NSF should plan on supporting activities that spur international collaboration around them.

Question 5: **How might NSF work with stakeholders to promote best practices for governance of research in emerging technologies at every stage of research?**

Summary: Anticipatory governance is itself a process that includes stakeholder visions, and one participatory Technology Assessment is an important component of doing so.

Anticipatory governance is precisely a vision for doing this through the research process. Thus, NSF should socialize the vision of anticipatory governance (and its component capacities of anticipation, engagement and integration) and that of cognate visions around which there has been significant intellectual development (e.g., responsible innovation, public interest technology, and participatory Technology Assessment) with its broader stakeholders. Such activity would highlight that these techniques are neither "anti-science" nor "anti-technology;" nor are they out of place in many high-tech endeavors in the private sector. NSF should also conduct or provide for the conduct of a review of its own sponsored research on such best practices. The long history of NSF's earlier Ethics and Value Studies program, the more recent history of the STS program and the SciSIP program, the Centers for Nanotechnology in Society (at ASU and UC Santa Barbara), all provide relevant experience for such a review.

***Question 6:* How could ethical, social, safety, and/or security considerations be incorporated into the instructions for proposers or into NSF's merit review process? Also, what challenges could arise if the merit review process is modified to include such considerations?**

Summary: Two logical alternatives for including ES3 in review are 1) making it part of intellectual merit; and 2) making it an addendum akin to the post-doc mentoring plan or data management plan. It is also important for NSF to consider how to ensure quality peer review of ES3 plans.

Since the CHIPS Act specifically excludes the possibility of expanding the existing "broader impacts" category to include ES3 topics, but also requires that these topics be subject to peer review, there seems to be two logical alternatives: 1) to include ES3 topics in the intellectual merit category, or 2) to create an addendum to each proposal akin to the data management plan or the post-doc mentoring plan. While I believe the former is more compelling and rigorous intellectually, I also believe it would be more politically challenging among NSF's constituency of (prospective) investigators, and that the second option makes more sense operationally.

One foreseeable challenge, however, of the second option is that institutions will develop cookie-cutter responses for such addenda, and that solid and influential intellectual work would not follow. It might be reasonable for some templated response at the institutional level to exist, e.g., my institution (Arizona State University) has a university-wide "design aspiration" articulated by the central administration that we faculty should "practice principled innovation." Somewhat similarly, Howard University's motto is "*Veritas et Utilitas*" (Truth and Service). It would be problematic to forbid or penalize institutions for mobilizing the intellectual resources of such related efforts, but at the same time much of the information submitted and reviewed really should be project-specific rather than institution-specific.
To get the best intellectual result out of the requirement, as well as to have it influence the direction and content of the research most directly and appropriately, it is crucial that ES3 work be understood as already integrated with the proposal itself as well as with the proposed work. I have seen poor examples of integration, e.g., serving as an external reviewer on an Engineering Research Center (ERC) in a case in which a social science component was required of the technical work as a condition of the award, rather than arising from the vision of the proposal. Predictable difficulties followed, e.g., lack of shared vocabulary, lack of repect, and ultimately alienation of the research teams from each other.

NSF also is regarded by many among its external constituents as having difficulty conducting peer reviews in new and, particularly, interdisciplinary programs. It is difficult to find the right reviewers or correct (substantive and other) balance of reviewers for such new programs. The difficulty here would of course be multiplied by the sheer number of programs involved across all of NSF. At risk of delaying the full implementation of the CHIPS ES3 mandate, NSF might consider pilot implementation through a certain number of its programs, perhaps those that are more interdisciplinarily inclined in the first place, or those of largest scale to begin with. Such staging on NSF's part would not necessarily relieve issues of compliance on the part of applicant institutions, although a relatively smaller subset of institutions take part in Regional Innovation Engine, Science and Technology Center, infrastructure, and ERC competitions than do across the full scope of program areas.

Question 7: **What other measures could NSF consider as it seeks to identify and mitigate ethical, social, safety, and/or security risks from research projects or other activities that the agency supports?**

Summary: NSF could model its approach to ES3 on other existing programs and solicitations of the broader community, e.g., through planning grants and workshops prepare for larger-scale research. It is crucial for NSF to draw on the significant record that its funding has already produced in this area.

Again drawing on the CHIPS Act list of technologies, NSF could fund and organize an ES3 workshop or planning grant competition for each technology to begin to explore and map out the issues that could then become the core of more extended research calls. This approach would be superior to, say, contracting with the National Academy of Sciences, Engineering and Medicine on such topics because it would be closer to NSF's constituencies across many disciplines. Analogous to how the current CRISES competition is structured, for example, NSF could solicit competitive proposals for a workshop and/or planning grant for the ultimate development of a center-level proposal for each of the CHIPS technology areas.

At a smaller scale, NSF could also create specific, embedded scholar programs at the graduate, post-doc, and faculty level that enabled ES3 scholars to be interact with STEM research groups or centers, and vice-versa. Building on another model developed at ASU that drew from the embedded scholar ("socio-technical integration research" or STIR) work, NSF could fund scholar programs in the "STIR City" model (Richter et al. 2017; NSF#1535120) in which a societal researcher followed a particular technology or innovation across multiple sectors within an innovation ecology, in order to get a fuller understanding of the societal issues embedded in academic, private sector, and public sector institutions. Such STIR activities could begin relatively quickly and easily if NSF were to solicit proposals for supplements to existing medium- and large-scale funded activities (with a social scientist or relevant humanist as named co-PI).

Ultimately, NSF should take the work that it has funded over time through is various social science, ethics and values studies, science and technology studies, ethical and responsible research, and responsible development programs seriously and build on their successes.