

## Vision for a 4D Center

As we think more concretely about a possible initiative in the realm of applied data science it is critical to articulate the answers to three questions: What might be new and different about a “4D Center”? What would that center do? And what are realistic metrics of success?

Perhaps the most important idea is that what we’re proposing is not a traditional “data science center,” a number of which exist or have been proposed at universities and institutions. Data science centers tend to be service arms of a research organization, with teams of data scientists waiting in their facility to provide discipline scientists with access to an arsenal of analytical and visualization capabilities. That scenario is common and can be extremely useful, just as machine shops or libraries are useful in advancing research.

However, what we envision as the 4D Center is radically different—a *science discovery center*, with the motivation and agenda driven by big science questions in the realm of the Carnegie Institution’s great strengths—Earth, space, and life. Consequently, the organization and activities of the 4D Center will be different from a traditional data science center. Discipline scientists from Carnegie and numerous collaborating institutions will take the lead, employing data science methods to foster discovery through cross-disciplinary, multidimensional strategies.

Our thinking in how to structure and advance such a unique endeavor is informed by the success of the Deep Carbon Observatory (DCO). We see the 4D Center advancing in three sequential stages, each dependent on increasing levels of fundraising.

*Step 1—Advance Data-Driven Discovery within the Nascent 4D Community:* The 4D Center would initially serve to integrate and advance data-driven discovery by engaging and coordinating the core network of more than 100 active 4D Workshop participants and other current collaborators in Earth, space, life, and data science—eager researchers from more than 20 institutions in a dozen countries. We will identify key unanswered questions that are particularly amenable to multidimensional, multivariate strategies—in particular questions related to planetary evolution, with special emphasis on the co-evolution of the geosphere and biosphere on Earth. We will build and sustain relevant data resources, link those resources across domains, and enlist powerful advanced analytical and visualization methods to tease out previously unrecognized patterns. The 4D Center will thus work with an international team on scientific challenges of interest to a broad range of scientists, leveraging the interests and expertise of existing collaborators to accelerate data-driven discovery.

A unique aspect of this concept is the pursuit of cross-disciplinary integration by applying multidimensional, multivariate advanced analytics to accelerate discovery in our traditional domains. At Carnegie, with our flexible vision and broad scientific interests, we are uniquely positioned to embrace and lead this opportunity for tackling problems that lie at the intersection of multiple fields of science. An additional opportunity lies in training early-career scientists, who will become proficient at coding and using data science in their domains. So trained, they will have great advantages in subsequent stages of their careers. Scientists who understand how to manipulate and explore data in higher dimensions—to discern overlooked, hidden trends—will be poised to make great discoveries in the coming decades. The 4D Center can engage in the active education and training for those scientists who want to learn these approaches.

This first step in the evolution of the 4D Center can be accomplished with a minimal initial investment, as there is no new infrastructure required. The core network of 4D Center scientists, including a team of early-career researchers, can make tremendous progress. Modest resources of \$300,000 per year in external funding could thus support a dynamic program.

*Step 2. Expanding and Guiding an International Consortium:* A second step in the evolution of the 4D Center, much in the spirit of the DCO, is to significantly expand the international consortium of Earth, space, life, and data scientists who share this vision and are dedicated to advancing the goal of data-driven discovery, thereby achieving tremendous leveraging of our initial investment. To accomplish this second ambition we would require perhaps \$1 M per year in grant support, primarily to run a small office, hold annual meetings and workshops, and provide targeted funding to encourage promising individual initiatives around the world.

As with the DCO, leveraging an investment is the key. Modest amounts of funding can be enhanced by more than an order of magnitude if thoughtfully distributed. For example, DCO distributed dozens of small awards of \$5,000 to \$10,000 to pay summer salaries of researchers to write grants—efforts leading to many millions of dollars in new funding around the world. Funds allocated to support targeted workshops and special sessions at conferences will provide catalysts for exciting new collaborations and grants. Participating DCO scientists have benefitted tremendously from such team-building experiences, and they will collectively become the world leaders in a dynamic emerging field.

*Step 3—Training the Next Generation of Scientists:* Ultimately, a compelling metric for success is the creation of an expanding web of institutional partners that ultimately results in training hundreds of early-career Earth, space, and life scientists in the philosophy and methods of data-driven discovery. We would begin by finding external support to develop a web-based “4D Curriculum” for graduate students and postdocs, focusing initial training efforts at perhaps a dozen institutions and intensive 4D Training Workshops. As those young scientists gain confidence in the methods of data science, and as they move on to begin their own faculty positions, the training network will expand. That international “4D network” of early-career scientists, once established and thriving, will be a tremendous legacy and tribute to the funder’s vision.

This third step could employ significant financial support, including stipends for early-career scientists, support for their home institutions, significant travel funds to 4D workshops, training events, and conferences around the globe, and other expenses associated with a dynamic and thriving scientific enterprise. Funding objectives for Step 3 must remain open-ended, but an aspirational budget approaching \$10 M per year for 10 years is not out of reach.

The 4D Center remains a vision, an aspiration. Nothing like it now exists. But the opportunity is clear and compelling. The initial investment is modest, while the potential rewards are great—the opportunities are not just to accelerate the pace of discovery in Earth, space, and life sciences, but to transform our understanding of the cosmos through a new, vivid, multidimensional perspective.