Honorable Gina Raimondo
Secretary
U.S. Department of Commerce
1401 Constitution Avenue NW
Washington, D.C. 20230

Re: AI-Ready Open Data Assets RFI, DOC-2024-0007

Dear Secretary Raimondo:

Benchmark Labs is a San Diego, California-based startup that provides farmers with actionable weather data relevant to their location to help them save water, energy, pesticides, and fertilizers while increasing crop yield. Consumer weather models are too coarse for farms and land conservancies that have their own microclimates, so we use machine learning to build custom, tailored weather models for farmers, land managers, and others. To build these models we rely on data from both public and private sector sources. We accordingly have great interest in the availability of data for AI applications, and appreciate the opportunity to provide input as the Commerce Department considers the development of AI-ready open data assets.

Weather has a significant impact on the economy, especially in agriculture. For example, pests and insects often have growing cycles that are related to climate variables like temperature and relative humidity. Labor scheduling, tilling, harvesting, spraying, fertilization, everything is driven by the weather. That is why we use technologies like machine learning to improve weather forecasting and deliver tailored insights to farmers.

Traditionally, the National Weather Service and third-party weather providers divide the world into geographical “boxes.” These tend to be too broad since everybody inside of a box receives the same forecast. While this approach works for many applications, like understanding the movement of a tropical hurricane in the Atlantic, if you are a farm owner, you first care about what is important for your farm. After that comes what is happening at the county-level and then the state-level.

Benchmark Labs refines the forecasts from the National Weather Service -and other weather providers that divide the world in boxes- by using information farm specific information and our proprietary machine-learning techniques. We also use cloud-computing services to train machine learning models that refine those core

models in real-time. Right now, we are producing about 5,000 models per week. In the end, we provide more actionable weather forecasts, recommendations, and alerts that are tailored to a specific location.

The government data we use today is made available via cloud providers (like AWS, Google Cloud, and Microsoft Azure) thanks to an open data initiative that began about 10 years ago. It is intuitive and cost efficient, but this data is only available with a lag caused by data handling and postprocessing. Real-time data is available directly from the National Weather Service, but it is made available through HTTP or FTP servers, which are costlier for us to query. (And I suspect that it is less accessible and intuitive for those without advanced degrees or experience).

By comparison, the European Centre for Medium-Range Weather Forecasts is more accessible and could be a model for the Commerce Department's current efforts. They have better documentation, and some data already have a license already assigned. That clarity lowers practical barriers to data use, and should be emulated. The Europeans also have a dedicated data portal, where portal users can select the data they want, query it by volume, or access by API.

Easier access to government data should exist on a spectrum that meets a range of potential data users at their skill levels. For example, you can imagine students making use of a dedicated data portal to download specific data in an excel spreadsheet, while more sophisticated researchers, financial institutions or commercial entities may use an API. Moreover, if accessibility is viewed on a spectrum, then data quality and utility need not be compromised to further access. That spectrum should include offering data as raw as possible, including in obscure formats, if necessary. Access to this raw data will enable research and innovation in ways that might not be possible with pre-refined datasets.

Finally, the Department of Commerce should make available AI-ready data from all its Bureaus (or as many possible), not just those that are typically thought of as data producers like the Bureau of Economic Analysis or Census Bureau. For example, a startup could use International Trade Administration data to help organizations understand supply chains. A startup support organization could use Economic Development Administration data to understand where to direct resources. And in our case, we make use of National Oceanic and Atmospheric Administration data to provide farmers with tailored forecasts. Making AI-ready data resources available from across the Department is essential and will lead to innovative solutions that address issues that span the Department's broad jurisdiction.

AI is based on statistics, and to have robust statistics, you need data. It's a numbers game, and for startups like ours to be able to compete, we need to access this data. There are a lot of opportunities for smaller companies to fill service gaps and innovate in areas that may not be of importance to larger entities that serve broader missions. Benchmark Labs is an example of that. Thanks to the data we use from the National Weather Service we fulfill a need that they cannot, because providing such localized information is not within their scope or purpose. Our experience shows both the value of open data and the need for further improvements in the availability of government data and AI-ready data assets. We appreciate the Department's interest in exploring these improvements and are grateful for the opportunity to share our perspective.

Sincerely,

Dr. Carlos F. Gaitan Ospina, Ph.D.
Co-Founder and CEO
Benchmark Labs, Inc.
4660 La Jolla Village Dr.
San Diego, CA 92122