DRAFT US CLIVAR WHITE PAPER

SUMMARIZING WEATHER, CLIMATE, AND EARTH SYSTEM OBSERVATIONAL DATA SHARING NEEDS FOR RESEARCH AND EDUCATION

Authors: Yolande Serra[1], Carol Anne Clayson[2], Mike Patterson[3], Aneesh Subramanian[4]

INTRODUCTION

Free, open, and unrestricted access to and sharing of atmospheric, oceanic, and other Earth system observational data is a foundational principle, reflected in long-standing international and U.S. data policies, that underpins and enables research and education in weather, climate, and Earth system sciences. The U.S. government and scientific community have historically joined with partners internationally to champion the establishment and continuity of such data sharing policies to advance monitoring, understanding, and prediction of weather, climate, and the Earth system, enable reproducibility of research findings, and support the training of new generations of scientists.

Scientific research that leads to improved weather and climate information for societal benefit is often limited by data scarcity. Increases in both data volume and variety from an ever-growing number of public and private sources can help fuel innovation and scientific understanding, but only if these data are free and openly accessible to the research and educational communities in order to maximize its value.

Preparations are underway for the fall 2021 Congress of the World Meteorological Organization (WMO), where an updated WMO Data Policy – reflecting an expanded suite of observational data, the role of private commercialization efforts, and other drivers – will be discussed and potentially adopted. This meeting provides a critical opportunity for the U.S. government representation to reaffirm our nation's continued commitment to free, open, and unrestricted observational data access and sharing.

Within the U.S., federal agencies have individually adopted data policies that adhere to the international agreements for free, open, and unrestricted data access and sharing. These include the National Oceanic and Atmospheric Administration (NOAA) and the National Aeronautics and Space Administration (NASA), with their mission responsibilities for acquiring, archiving, and providing access to and enable sharing of Earth system observational data.

During the past five years, NOAA and NASA have undertaken pilot programs to explore commercially sourced data buys and their efficacy in meeting agency-mission objectives. For some pilots, not all, the impact of data sharing restrictions on scientific research has been evaluated. As the agencies begin to move from pilot programs to establish ongoing data buys, it

[1] *University of Washington*
[2] *Woods Hole Oceanographic Institution*
[3] *US CLIVAR Project Office*
[4] *University of Colorado*

is our view that the impact of restrictions on data access and sharing provisions of awarded contracts on scientific research and education be assessed. Benefits accrue to the scientific endeavor when agencies uphold their stated policies for free, open, and unrestricted data access and exchange.

The primary purpose of this white paper is to articulate the needs and benefits of free, open, and unrestricted observational data access and exchange for scientific research and education purposes. The intent is two-fold – to inform future data policies in the U.S. and internationally as well as to inform future U.S. agency approaches to purchasing observational data. The document is presented in four sections: (i) examples of prior and recent U.S. agency commercial data buys and their evaluated and potential impacts; (ii) overview of international and U.S. organizational data policies of relevance; (iii) enumeration of the specific needs for data access and sharing to enable scientific research and education; and (iv) articulation of a clear, concise summary position statement. To clarify, we underscore that the intention is to inform, not advise, as US CLIVAR does not have an advisory role.

This white paper has been developed by a writing team charged by the US CLIVAR Scientific Steering Committee and incorporates inputs received through consultation with members of the US CLIVAR organization as well as feedback from members of the scientific community collected through two Town Halls and an open review period of the draft document. There may be additional views beyond what we have captured through our information collection efforts, and we understand and welcome that other organizations representing U.S. and international scientific community interests may offer their views as well.

BACKGROUND

In 2016 the U.S. Congress initiated the first Commercial Weather Data Pilot (CWDP) that directed the National Oceanic and Atmospheric Administration (NOAA) to purchase commercial weather data and evaluate its usefulness for their weather modeling and forecasting activities (184 Cong. Rec. H9736, 2015). The following year NOAA decided to explore commercial options for space-based Global Navigation Satellite Systems (GNSS) radio occultation (RO) data, which can easily be collected from smallsat platforms. In September 2017 NOAA awarded its first commercial data buy for RO data. Over $1M was spent in the acquisition of these data.

In April 2017 the Weather Research and Forecasting Innovation Act (2017) was signed into law requiring that NOAA evaluate the feasibility of operational data buys as an alternative to NOAA's space program for weather data. The Act upheld the U.S. commitment to the WMO's Resolution 40 (WMO 1995; Bautista Pérez 1996, WMO 2021a), signed by member-states in 1995, which holds its members to open data sharing of essential meteorological parameters. At the time Resolution 40 was signed, however, satellite data was not clearly covered under the essential data clause. The Weather Act covers both satellite data and hosted payloads, requiring NOAA to justify public sector development of systems to collect such data over the commercial sector going forward.

In September 2018 NOAA awarded its second round of RO commercial data buys under the CWDP program, amounting to over $8M. After determining that commercial RO data are acceptable for operational use, NOAA awarded its first commercial RO contracts in support of

operational weather forecasting in November 2020. NOAA has requested $23M in support of this two-year contract with commercial RO data providers, as well as for additional pilot studies of new commercial space technologies beyond RO. And while the Weather Act of 2017 assures the U.S. commitment to WMO Resolution 40, RO data have only been around since the early 2000s and thus do not clearly come under the data sharing provisions of Resolution 40. The new commercial contracts to purchase RO data are restricted to U.S. government use within the first 24 hours, after which these data can be shared openly without restriction (Griffin et al. 2021).

RO data are not the first data NOAA has purchased from commercial vendors. In September 2016 NOAA contracted Earth Networks Inc to provide ground-based Global Positioning System Meteorology (GPS-Met) total column water observations. These data are not allowed to be distributed by NOAA, nor are they allowed to be publicly distributed on the Global Transmission System (GTS) used to provide data around the world for weather forecasting (NOAA 2021). The contract with Earth Networks Inc will come to an end in summer 2021, and a new Request for Proposals to provide GPS-Met data will be released in early 2021. This provides an opportunity for NOAA to renegotiate the contract on these data to be less restrictive, putting these data back in the hands of the research and education community and thus enhancing the overall value of these data to society.

The commercialization of ground-based lightning data occurred even earlier. In the 1980s NOAA was working on an imager for their geostationary satellite and deciding whether they would develop their own ground-based network or purchase these data from Vaisala. In the end, a data buy was established with Vaisala without the option to distribute the data outside of NOAA. Vaisala does offer its data for purchase through a grant writing process .

NOAA is not the only U.S. agency to procure commercial data. In December 2017 NASA launched the Private-Sector Small Constellation Satellite Data Product Pilot or Commercial Smallsat Data Acquisition Program (CSDAP) (NASA 2021). For the first Request for Information (RFI) of this program, four contracts were awarded to commercial vendors for high spatial resolution multispectral imagery and GNSS related products. Data purchased under this program are archived at NASA and can be made available only to current and future NASA-funded researchers (NASA Earth Science Division 2020). Going forward, every 12-18 months NASA will release new RFIs for commercial data of potential use towards NASA's research and applications objectives. According to NASA's own evaluation of the first round of data buys associated with the program (NASA Earth Science Division 2020), the strictness of the End User License Agreements (EULAs) was a significant hindrance to the usefulness of the commercial data in the advancement of science. In addition, NASA found that the EULAs "created additional complexity and cost" for NASA. An evaluation of products by all six Earth Science Division thematic research areas and four applied science programs determined that data sharing licensing terms significantly hinder scientific collaboration. Specific concerns identified include (i) meeting publication requirements for sharing source data for reproducibility, (ii) the distribution of derivatives, such as incorporation into atmospheric reanalysis, and (iii) sharing of imagery and data across the US federal agencies and partners. Despite these findings, NASA has decided to proceed with similar EULAs for their commercial data buys, concluding that the usefulness of

the data for NASA science goals outweighed the issues surrounding the EULAs in contradiction to the agency's own stated policies.

Given these recent examples of commercial data buys, as well as many more outlined in a report of the National Research Council (NRC 2001), the research and education community needs to take notice of commercial data buys by NOAA and NASA and other U.S. agencies. Strict EULAs on weather and climate datasets will, by NASA's own findings (NASA Earth Science Division 2020), inhibit the free flow of information and thus scientific progress. The U.S. has been one of the biggest advocates in providing free and open access weather and climate data. And just as others have begun to follow this example, the U.S. is now choosing to restrict access to global data that could reasonably be said to contribute "to the protection of life and property and the well-being of all nations" – WMO Resolution 40, 1996.

CURRENT DATA POLICIES

The WMO Resolution 40 (WMO 1995; Bautista Pérez, M., 1996; WMO 2021a), signed by WMO members including the U.S. in 1995, "established agreed upon policy and practices for the exchange of meteorological and related data and products" including commercial activities. This resolution committed its members to providing timely, free, and open access to all meteorological and related data products that served in "the protection of life and property and the well-being of all nations." This commitment to free, open and timely access to meteorological data was expanded to include hydrological data (WMO 1999; WMO 2021b) and climate data (WMO 2015; WMO 2021c) in 1999 and 2015, respectively, in recognition of the importance of the availability of data to all earth system science research and education efforts. However, since the writing of WMO Resolutions 40, 25, and 60, there have been additional advancements in the type of observational data collected as well as an expansion of the private sector involvement in procuring these data. As a result, the WMO is updating these Resolutions with Resolution 42 (WMO 2021d), which, if accepted, will be ratified at the WMO Executive Council and the World Meteorological Congress in 2021. This updated resolution, which WMO refers to as the "Unified Data Policy" document, reaffirms the requirement that members provide free, timely, and open access to weather and climate data, and expands the type of data covered under the agreement to earth-system data as a whole. The resolution also expands the data covered under the resolution to include data from both the public and private sectors. The updated resolution explicitly calls out the requirement for free and unrestricted earth-system data to "the research and education communities, for their non-commercial activities" (WMO 2021d).

Following the signing of the WMO Resolution 40, the Intergovernmental Oceanographic Commission (IOC) adopted Resolution IOC-XXII-6 "IOC Oceanographic Data Exchange Policy" (IOC 2003). This policy commits its member states to provide timely, free and unrestricted access to oceanographic data and associated metadata and products generated under IOC programs. The IOC member states decided that these data are vital to operational weather, marine environment and climate prediction, the preservation of life and property, the mitigation of human-induced changes to the marine and coastal environment, and for the advancement of scientific understanding.

The International Science Council (ISC) includes representatives from the International Council for Science, the InterAcademy Partnership, The World Academy of Sciences, and the International Social Science Council, which together represent the broad interests of the global scientific community in international policy for science. The ISC argues in a 2017 statement on "Open Data in a Big Data World" that open data is "critical to assure the rigor of research findings" by proving the access to these data to the global community needed to replicate important and potentially controversial results. In addition, they note the important equalizing impact on the least-developed countries by allowing them to participate more fully and equally in global research efforts (Science International 2015).

The open data policies endorsed by these international scientific organizations are echoed in the data policies of U.S. scientific societies including the American Meteorological Society (AMS) 2019 Full, Open and Timely Access to Data Statement (AMS 2019), an update to a similar statement released by the AMS in 2013 (AMS 2013). The American Geophysical Union (AGU), representing the broader geoscience community, has also had a position statement since 1993 on the need for scientific data to be widely available for archiving purposes, to support and advance scientific exploration and discovery, and to guarantee the integrity of scientific results through independent evaluation and replication by the global community (AGU 2019). In their commitment to their open data policy, they require authors to provide access to all data and code used to generate results published in their journals.

Some outstanding issues remain in finalizing WMO Resolution 42. In particular, clarifying who the policy applies to and what is the obligation and encouragement of stakeholders beyond national meteorological services. There also remains the issue of clearly defining what is 'essential' in a practical and meaningful way. The metric that defines essential variables for international observing systems (e.g., GCOS, GOOS) should not necessarily be adopted for this purpose. For scientific purposes users need access to all readiness levels of data products, not just those that are well established and already used for weather, climate, hydrological and oceanographic forecasting. The new document must also be flexible enough to reflect ongoing changes within the global community and the changing landscape of weather, water, climate and related data so that it does not become quickly irrelevant. As noted during the WMO Data Conference in November 2020 (WMO 2020), Resolution 42 cannot be prescriptive, but must guide, encourage, and incentivize data sharing as well as new players in the arena of data collection.

The collection and access to high-quality earth-system data, as well as its long-term preservation, is not guaranteed. Scientists and educators need to stay vigilant and engaged in the policy decisions affecting the integrity, exploration and advancement of scientific endeavors. While some shorter term objectives behind the collection of earth-science data may be served by more restricted data access, the broader goals of advancing earth-system science for the betterment of global society, transparency in the scientific process, and educating the next generation of earth-system scientists are not served by restricted access to or limits on sharing of earth-system data or the cessation of collecting a subset of these data due to a lack of commercial value or vendors.

Data Requirements for Research and Education

Researchers and educators are important stakeholders in the collection, use, and archiving of scientific data. They comprise the community that is responsible for advancing scientific knowledge and training the scientists and application specialists of tomorrow. At the same time, this community is dispersed among universities, government laboratories, and private sectorresearch groups around the world, making it difficult to organize and advocate its needs to those responsible for providing the data needed to explore and advance weather, climate, and Earth system science and educate the next generation. Much of the work in defining the needs of the research and education community regarding scientific data policy has been done by both U.S. professional societies representing these communities (e.g., AGU, AMS), as well as global organizations tasked with designing policy related to scientific data for the protection and betterment of global society (e.g., WMO, IOC, ISC). Those policies, reviewed in the previous section, have a common theme, that free and open access to weather, climate, and Earth system science data is foundational for advancing science in support of society, maintaining scientific integrity and educating the next generation of scientists. Drawing from the policy statements of these organizations noted in the previous section, as well as analyses done in establishment of these policies, this section summarizes the key requirements of the research and education community regarding access, quality expectations and archiving of scientific data.

Requirement 1: Timely and unrestricted access to quality weather, climate, and Earth system science data and the long-term archival of the raw data are essential in order for science to best serve society.

Scientific progress relies on free flowing information and ideas. According to the NRC report (2001), scientists are opportunistic and restrictions on data can result in scientists abandoning those lines of research whose data have burdensome restrictions. Unrestricted access to raw data additionally supports new uses of those data as computational resources and fundamental understanding evolve (e.g., NRC 2001; Zillman 2020). Establishing requirements regarding the needed accuracy and precision of data is an important aspect of any data provider contract agreement. Archiving policies are also important to climate science, which requires long time records of global essential variables in order to validate model projections and monitor important processes affecting the Earth system over time. Regardless of who collects observational weather, climate, and Earth system science data, these data have the greatest value to society when used by the most scientists and for the greatest number of applications. This is best accomplished by collecting high quality data with minimal restrictions and cost to the users and maintaining an archive of those data for future use.

Requirement 2: Timely and unrestricted access to all weather, climate, and Earth system science data at little to no cost to encourage equal opportunity around the globe, including contributions from least-developed nations, in scientific exploration and discovery.

Commercial interests focused within one country may not support the global collection of data required by the research and education communities in pursuit of understanding and exploring the global earth system (e.g., NRC 2001; GEO 2015; Zillman 2020). Governments that purchase weather, climate, and Earth system science data and agree to share these data equally and freely assure a greater diversity of perspectives in interpreting those data, helping support the best

scientific outcomes for societies around the world. Unrestricted access to data is the best assurance for equity and diversity both in terms of the scientists who do the science, as well as in those who benefit from the outcome of the research.

Requirement 3: Timely and unrestricted access to all weather, climate, and Earth system science data and the long-term archival of the raw data are essential for maintaining the integrity of scientific results through unhindered access to verify and replicate such results by scientists around the world.

Most reputable scientific professional societies require that scientists who publish in their journals make their data and methods publicly available so that their results can be verified and reproduced. Ultimately, scientific knowledge is based on results that have been reproduced many times under a broad range of circumstances and by many different scientists with different backgrounds and perspectives. This reproducibility builds scientific integrity and public trust in the use of science for the public good.

Requirement 4: Timely and unrestricted access to all weather, climate, and Earth science data to educate the next generation of the operational and investigative work force in the field of weather, climate, and Earth system science.

Educators access both archived and real time data for training the next generation of weather, climate, and Earth science professionals. Free and low cost access to observations makes it possible for students and educators from many different backgrounds to have equal access to the information they need to build careers in weather, climate, or Earth system science.

The above needs are based on decades of consideration regarding best practices in support of scientific exploration and discovery, including preserving equal participation in the global scientific process by all nations and the need for preserving the observations used as a basis for the current knowledge and understanding of weather, climate, and Earth system science. In addition, in order to advance understanding of earth's climate, high-quality data need to be collected and archived over long time periods. The only exception to the requirement for free, open, and timely access to Earth system data is if those data are protected by public disclosure laws, as is often the case in the area of public health and national security data sets. In such cases, it remains the responsibility of the U.S. government to assure the advancement of knowledge for the public good and the integrity of the scientific process, while also adhering to the laws protecting the data.

SUMMARY

Ultimately, the research and education communities in weather, climate, and Earth system science are agnostic to the provider of the data collected in support of their science. Data, whether originating from public institutions, private institutions, or public-private partnerships, are the foundation supporting advancements in weather, climate, and Earth system science and, ultimately, the benefits of those advancements to society. Scientific and educational pursuits provide the most benefit to society when these data are provided to the global scientific community through a timely, free or low cost, and unrestricted process. Data policies that consider only the costs of a particular agency or institution may in the short term result in a lower

cost end user license agreement, however, such thinking is short sighted and will ultimately cause those agencies or institutions to lose ground on their longer term objectives of advancing scientific knowledge and understanding in the earth-system sciences to the benefit of the societies to which they serve. In anticipation of WMO 42 and considering further unforeseen advancements in data collection, the U.S. government has a responsibility to maintain its commitment to global data policies and to maintain consistency in this commitment over time and across agencies.

In putting this report together, several important issues were raised that we decided were outside of the scope of this White Paper but still deserved mention, as they are intricately linked to the data sharing issue. In particular, the community notes the strong need not only for free and open access to data in the Earth system sciences, but also access to free or low-cost tools for storing, accessing, and analyzing these data, particularly within developing countries. Also mentioned was the need for training on the use of such tools. These important issues should be addressed in order to truly achieve the four requirements for research and education listed in this White Paper. Without equalizing opportunity around the globe with respect to data access, including tools to manipulate these data, we will not have full participation in the scientific processes and thus risk missing important discoveries.

Another important issue that arose in our discussions with the research and commercial sectors regarding scientific data is that of data delivery. The scope of this White Paper is limited to arguing for the need for free and open access to weather, climate, and Earth system data, but does not take up the issue of how to get these data to scientists and educators around the world. The commercial sector has played a significant role in providing the systems that can deliver large amounts of data to a variety of stakeholders that rely on these data. Public-private partnerships of this nature also come with their own set of licensing agreements that need to be vetted to assure the needs of the research and educational communities are met equitably around the world.

References

184 Cong. Rec. H9736 (daily ed. Dec. 17, 2015), https://www.congress.gov/crec/2015/12/17/CREC-2015-12-17-house-bk2.pdf.

American Geophysical Union, 2016: AGU Data Policy, Accessed 22 February 2021, https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Policies/Data-policy.

American Meteorological Society, 2013: Full and Open Access to Date—A policy statement of the American Meteorological Society, Accessed 22 February 2021, https://www.ametsoc.org/index.cfm/ams/about-ams/ams-statements/archive-statements-of-the-ams/full-and-open-access-to-data/.

American Meteorological Society, 2019: Full and Open Access to Date—A policy statement of the American Meteorological Society, Accessed 22 February 2021, https://www.ametsoc.org/index.cfm/ams/about-ams/ams-statements/statements-of-the-ams-in-force/full-open-and-timely-access-to-data/.

Bautista Pérez, M., 1996: Resolution 40 (Cg-XII)—WMO policy and practice for the exchange of meteorological and related data and products, including guidelines on relationships in commercial meteorological activities. *WMO Bulletin*, 45(1):24–29, https://www.wmo.int/pages/prog/www/ois/Operational_Information/Publications/Congress/Cg_XII/res40_en.html.

Griffin, V. L., D. Spencer, F. Gallagher III, P. C. Sullivan, E. C. Grigsby, G. Mandt, M. Scott, E. Talaat, S. Jacobs, D. Marks, and G. Bravo 2021: Overview of future architecture implementation as informed by the NOAA 2019 Broad Agency Announcements. *7th Annual Symposium on Operational Environmental Satellite Systems,* Amer. Meteor. Soc., 5.1, https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/379751.

Group on Earth Observations, 2015: *The Value of Open Data Sharing*. GEO Report, 42pp, https://www.earthobservations.org/documents/dsp/20151130_the_value_of_open_data_sharing.pdf.

Intergovernmental Oceanographic Commission, 2003: *Twenty-second Session of the Assembly*. IOC Report SC.2003/CONF.214/CLD.37, 164pp, https://www.jodc.go.jp/info/ioc_doc/Governing/131400e.pdf.

NASA, 2021: Commercial Smallsat Data Acquisition (CSDA) Program. Accessed 22 February 2021, https://earthdata.nasa.gov/esds/csdap.

NASA Earth Science Division, 2020: *Commercial SmallSat Data Acquisition Program Pilot Evaluation Report*. 24pp, https://cdn.earthdata.nasa.gov/conduit/upload/14180/CSDAPEvaluationReport_Apr20.pdf.

NOAA, 2021: GPS-Met Dataset. Accessed 22 February 2021, https://madis.ncep.noaa.gov/madis_gpsmet.shtml.

National Research Council, 2001: *Resolving Conflicts Arising from the Privatization of Environmental Data*. Washington, DC. The National Academies Press, 113pp, https://doi.org/10.17226/10237.

Science International, 2015: *Open Data in a Big Data World: An International Accord*. International Council for Science (ICSU), International Social Science Council (ISSC), The World Academy of Sciences (TWAS), InterAcademy Partnership (IAP), 16pp, https://council.science/wp-content/uploads/2017/04/open-data-in-big-data-world_long.pdf.

Weather Research and Forecasting Innovation Act of 2017, Pub. L. No. 115-25, 131 Stat. 91 (2017). https://www.congress.gov/bill/115th-congress/house-bill/353/text#toc-HF5628EAB6E1F4B22B0739D58D6A32508

World Meteorological Organization, 1995: *Twelfth World Meteorological Congress: Abridged Final Report with Resolutions*. WMO-No. 827, 144pp, https://library.wmo.int/doc_num.php?explnum_id=6033.

World Meteorological Organization, 1999: *Thirteenth World Meteorological Congress: Abridged Final Report with Resolutions*. WMO-No. 902, 171pp, https://library.wmo.int/doc_num.php?explnum_id=6019.

World Meteorological Organization, 2015: *Seventeenth World Meteorological Congress: Abridged Final Report with Resolutions*. WMO-No. 1157, 706pp, https://library.wmo.int/doc_num.php?explnum_id=3138.

World Meteorological Organization, 2020: WMO Data Conference—Summary and key points, Accessed 22 February 2021, https://meetings.wmo.int/WMO-Data-Conference/Documents/Conference%20summary%20statement.pdf.

World Meteorological Organization, 2021a: Resolution 40 (Cg-XII)—WMO policy and practice for the exchange of meteorological and related data and products including guidelines on relationships in commercial meteorological activities, Accessed 22 February 2021, https://www.wmo.int/pages/prog/www/ois/Operational_Information/Publications/Congress/Cg_XII/res40_en.html.

World Meteorological Organization, 2021b: Resolution 25 (Cg-XIII)—Exchange of hydrological data and products, Accessed 22 February 2021, https://www.wmo.int/pages/prog/hwrp/documents/Resolution_25.pdf.

World Meteorological Organization, 2021c: Resolution 60 (Cg-17)—WMO Policy for the international exchange of climate data and products to support the implementation of the global framework for climate services, Accessed 22 February 2021, https://library.wmo.int/doc_num.php?explnum_id=4192.

World Meteorological Organization, 2021d: WMO Data Policy Update, Accessed 22 February 2021, https://meetings.wmo.int/WMO-Data-Conference/Documents/Flyer%20for%20Res.42%2011%2015.pdf.

Zillman, J. W., 2020: Origin of WMO Data Exchange and Resolution 40, WMO Data Conference—Outcome Materials, Accessed 22 February 2021, https://meetings.wmo.int/WMO-Data-Conference/Documents/01_John%20Zillman_Origin%20of%20Data%20Exchange%20(Spoken%20version).pdf?Mobile=1.