

September 2007

CLIMATE CHANGE RESEARCH

Agencies Have
Data-Sharing Policies
but Could Do More to
Enhance the
Availability of Data
from Federally
Funded Research



Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE SEP 2007		2. REPORT TYPE		3. DATES COVERED 00-00-2007 to 00-00-2007	
4. TITLE AND SUBTITLE Climate Change Research. Agencies Have Data-Sharing Policies but Could Do More to Enhance the Availability of Data from Federally Funded Research				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Government Accountability Office, 441 G Street NW, Washington, DC, 20548				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 61	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			



Highlights of [GAO-07-1172](#), a report to congressional requesters

Why GAO Did This Study

Much of the nearly \$2 billion annual climate change research budget supports grants from the Department of Energy (DOE), National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), and National Science Foundation (NSF). Some of the data generated by this research are stored in online archives, but much remains in a less accessible format with individual researchers. As a result, some researchers are concerned about the availability of data.

GAO analyzed (1) the key issues that data-sharing policies should address; (2) the data-sharing requirements, policies, and practices for external climate change researchers funded by DOE, NASA, NOAA, and NSF; and (3) the extent to which these agencies foster data sharing. GAO examined requirements, policies, and practices and surveyed the 64 officials managing climate change grants at these agencies.

What GAO Recommends

GAO recommends the agencies explore opportunities in the grants process to better ensure the availability of data to other researchers and determine if additional archiving strategies are warranted. In commenting on a draft of this report, the four agencies generally agreed with our findings and recommendations. We incorporated technical clarifications as appropriate.

To view the full product, including the scope and methodology, click on [GAO-07-1172](#). For more information, contact John B. Stephenson at (202) 512-3841 or stephensonj@gao.gov.

CLIMATE CHANGE RESEARCH

Agencies Have Data-Sharing Policies but Could Do More to Enhance the Availability of Data from Federally Funded Research

What GAO Found

According to the scientific community—as represented by the National Academies and professional scientific associations—four key issues that data-sharing policies should address include what, how, and when data are to be shared, as well as the cost of making data available to other researchers. First, the information necessary to support major published results should be made available to other researchers. However, there are statutory limits on data sharing—such as intellectual property protections—as well as practical limits such as the lack of appropriate archives. Second, when the appropriate infrastructure exists, data should be made accessible through unrestricted archives. Third, data should generally be made available immediately or after a limited proprietary period to allow for analysis and publication of results. Fourth, data should be made available at no more than the marginal cost of reproduction and distribution. Finally, the extent to which specific policies address these key data-sharing issues may vary, depending on the type of research.

Although some program managers at all four agencies have included data-sharing requirements in grant awards, these agencies rely primarily on policies and practices to encourage researchers to make climate change data available. An interagency policy, as well as numerous agency, program, and project-specific data-sharing policies, encourages researchers to make climate change data available. The policies range from broad statements calling for open and timely access to data to more detailed policies that define the mechanisms and timelines for making the data accessible. Further, these policies often vary according to the needs of specific research programs or projects. Beyond their written requirements and policies, all of the agencies also rely on unwritten practices to facilitate data sharing. For example, two program managers withhold grant payments if data have not been made available for use by other researchers.

While the four agencies have taken steps to foster data sharing, they neither routinely monitor whether researchers make data available nor have fully overcome key obstacles and disincentives to data sharing. Because agencies do not monitor data sharing, they lack evidence on the extent to which researchers are making data available to others. Key obstacles and disincentives could also limit the availability of data. For example, one obstacle is the lack of archives for storing certain kinds of climate change data, such as some ecological data, which places a greater burden on the individual researcher to preserve it. Preparing data for future use is also a laborious and time-consuming task that can serve as a disincentive to data sharing. In addition, data preparation does not further a research career as does publishing results in journals. The scientific community generally rewards researchers who publish in journals, but preparation of data for others' use is not an important part of this reward structure. Consequently, researchers are less likely to focus on preserving data for future use, thereby putting the data at risk of being unavailable to other researchers.

Contents

Letter		1
	Results in Brief	3
	Background	5
	The Scientific Community Has Identified Several Key Issues That Policies Should Address to Facilitate Data Sharing	10
	Climate Change Research Agencies Rely on Various Policies and Practices to Encourage Researchers to Make Data Available	14
	All Four Agencies Have Taken Steps to Foster Data Sharing but Have Not Fully Overcome Key Obstacles	23
	Conclusions	26
	Recommendations	28
	Agency Comments and Our Evaluation	29
Appendix I	Objectives, Scope, and Methodology	31
Appendix II	Data-Sharing Policies Applicable to Federally Funded Climate Data	34
Appendix III	Examples of Data-Sharing Expectations for Different Federally Funded Research Projects	37
Appendix IV	Governmentwide Climate Change Data-Sharing Policy	39
	U.S. Global Change Research Program Data and Information Working Group Data Management for Global Change Research Policy Statements	39
	Background	39
	Applicability	40
	Guidelines and Their Application	40
	Suggested Data Product Requirement for Grants, Cooperative Agreements, and Contracts	46
	Compliance	46
Appendix V	Comments from the Department of Energy	47

Appendix VI	Comments from the National Aeronautics and Space Administration	50
Appendix VII	Comments from the Department of Commerce, for the National Oceanic and Atmospheric Administration	52
Appendix VIII	GAO Contact and Staff Acknowledgments	55
Figure	Figure 1: General Research Grant Process	8

Abbreviations

CCSM	Community Climate System Model
CCSP	Climate Change Science Program
DOE	Department of Energy
GLOBEC	Global Ocean Ecosystem Dynamics
MILAGRO	Megacity Initiative: Local and Global Research Observations
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NOAA	National Oceanic and Atmospheric Administration
NSF	National Science Foundation
OMB	Office of Management and Budget
PCMDI	Program for Climate Model Diagnosis and Intercomparison
REASoN	Research, Education and Applications Solution Network
RISA	Regional Integrated Sciences and Assessments

This is a work of the U.S. government and is not subject to copyright protection in the United States. The published product may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.



United States Government Accountability Office
Washington, DC 20548

September 28, 2007

The Honorable Joe Barton
Ranking Member
Committee on Energy and Commerce

The Honorable Ed Whitfield
Ranking Member
Subcommittee on Oversight and Investigations
Committee on Energy and Commerce
House of Representatives

The federal government invests nearly \$2 billion annually in climate change research, the majority of which is in the form of grants, cooperative agreements, and other awards funding researchers at external entities such as universities and privately owned research institutions.¹ Currently, electronic archives exist to systematically preserve some but not all kinds of data produced by federally funded climate change research. It is generally the responsibility of researchers to make data available, regardless of whether an appropriate archive exists for their use. Much of the data that are not made available through archives are retained by researchers and may be in a less accessible format. As a result, some researchers have expressed concerns about both the availability and long-term preservation of the rapidly growing body of climate change data.

According to the interagency Climate Change Science Program (CCSP), federal investment in interdisciplinary earth sciences, global observation systems, and satellite and computing technologies has improved our understanding of climate change. The CCSP coordinates and directs the climate change research performed by 13 departments and agencies. Four agencies in particular account for about 90 percent of the annual federal

¹The \$2 billion estimate is based on recent budget analyses from the U.S. Climate Change Science Program (CCSP), which compiles budget data from each agency that receives climate change funds. This estimate does not reflect all agency activities that support climate change science, such as operation of satellites and preservation of some kinds of climate change data. For more information on federal climate change spending, see GAO, *Climate Change: Federal Reports on Climate Change Funding Should Be Clearer and More Complete*, [GAO-05-461](#) (Washington, D.C.: Aug. 25, 2005).

climate change research budget.² The principal climate change research agencies—the Department of Energy (DOE), National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), and the National Science Foundation (NSF)—fund research through numerous programs.

Climate change data come from many specialized disciplines within the earth, physical, biological, engineering, and mathematical sciences. These data are often available in disparate forms, such as data output from models forecasting climate conditions, satellite images of ocean and land masses, and ice core samples from the Arctic region. For purposes of this report, we define data to include factual information or physical samples that are collected and recorded as a result of scientific observation, experiment, analysis, or similar methods of research. The output of models can also be considered data.³ The widespread availability of such data is important to developing a comprehensive understanding of climate change and its potential impacts. Indeed, committees of the National Academies, professional scientific associations, and federal research agencies regard the free and open exchange of data as an essential part of scientific research. A 1997 National Academies committee report stated that

“Governmental science agencies...should adopt as a fundamental operating principle the full and open exchange of scientific data. By ‘full and open exchange’ the committee means that the data and information derived from publicly funded research are made available with as few restrictions as possible, on a nondiscriminatory basis, for no more than the cost of reproduction and distribution.”⁴

²Based on the Fiscal Year 2006 estimated budget, the most recent year for which information about budget allocations were available.

³GAO is using the term *data* in this report to encompass both data and metadata. Metadata refers to information needed to understand the content, quality, and condition of the data, such as instrument calibration.

⁴National Research Council. *Bits of Power: Issues in Global Access to Scientific Data*. (Washington, D.C.: National Academy Press, 1997), p. 10. For additional National Academies’ reports on data sharing, see *Sharing Research Data* (1985); *A Question of Balance* (1999); *Resolving Conflicts Arising from the Privatization of Environmental Data* (2001); *Access to Research Data in the 21st Century* (2002); *The Role of Scientific and Technical Data and Information in the Public Domain* (2003); *Sharing Publication-Related Data and Materials* (2003); and *Expanding Access to Research Data* (2005).

In this context, you asked us to determine (1) the key issues that data-sharing policies should address as identified by the scientific community in order to facilitate the sharing of federally funded climate change data; (2) the requirements, policies, and practices for making data available to other researchers under current climate change research awards from the four major federal climate change research agencies; and (3) the extent to which the four agencies effectively foster data sharing.

In conducting our work, we identified and reviewed the data-sharing requirements, policies and practices that are part of climate change awards—primarily grants and cooperative agreements—funded by DOE, NASA, NOAA, and NSF. We also conducted a Web-based survey of the 64 program managers who oversee the climate change research awards at these agencies. We received a 100-percent response rate. We also interviewed senior officials at DOE, NASA, NOAA, and NSF who direct the climate change research programs as well as managers from data archives that preserve climate change data. Finally, we reviewed relevant data-sharing requirements, policies, and practices at other federal agencies, academic journals, and professional societies and conducted interviews with stakeholders representing those organizations (American Geophysical Union, American Meteorological Society, Ecological Society of America, the journal *Science*, the journal *Nature*, the National Academies, and the National Institutes of Health). A more detailed description of our scope and methodology is presented in appendix I. We performed our work between September 2006 and September 2007 in accordance with generally accepted government auditing standards.

Results in Brief

The scientific community as represented by the National Academies and professional scientific associations has identified several key issues that data-sharing policies should address, including what, how, and when data are to be shared, as well as the cost of making data available. The scientific community generally believes that, at a minimum, the information necessary to support researchers' major published results should be made available to other researchers. The scientific community, however, also acknowledges certain statutory limits on data sharing related to the protection of intellectual property, privacy, and national security, as well as practical limits to sharing, such as the lack of archival infrastructure. Nevertheless, it is generally accepted that when the appropriate infrastructure exists, data acquired in federally funded research should be made accessible through unrestricted archives. In terms of timing, the scientific community believes that data should generally be made available immediately or after a limited proprietary

period that allows researchers to complete their initial analysis and publish their results. The duration of such a period may be determined by the type of research. To address cost concerns, it is generally agreed that data should be made available at no charge or at least no more than the marginal cost of reproduction and distribution. Finally, the way in which specific policies address these key data-sharing issues may vary, depending on the type of research. Data-sharing policies must take into account their applicability to specific research projects, relevant legal and regulatory restrictions, the existence of appropriate archives, and the characteristics of particular research fields.

While some survey respondents at the four major climate change research agencies reported having incorporated data-sharing requirements into particular grant awards, each agency relies primarily on established policies and practices to encourage federally funded researchers to make data available. The policies we identified for all four agencies include the interagency CCSP data-sharing policy as well as agency, program, and project-specific policies that vary in how they address the key issues identified above. Agencies' policies range from broad statements calling for open and timely access to data to more detailed policies that define the mechanisms and timelines for making the data accessible. Further, we found that these policies vary among agencies and often vary according to the needs of a research program or project within the same agency. For example, the overarching data-sharing policy for NSF requires researchers to make data available to others but does not specify how, whereas the policy for NSF's ocean sciences program states that researchers should submit sediment samples from the ocean floor to particular archives for long-term preservation. We also found that large, collaborative research projects commonly have data-sharing policies unique to the project. For example, the AmeriFlux program—a network of climate change researchers funded by multiple agencies, including DOE, NASA, NOAA, and NSF—requires participants to submit data to a particular archive within 1 year of collection and specifies the preferred format for data submission. Beyond their written requirements and policies, all of the agencies also rely on unwritten practices to facilitate data sharing. For example, a majority of program managers surveyed identified archiving as one way for researchers to make data available. In addition, two program managers reported that they withhold installments of grant funds if researchers do not make data available. The use of such practices varies among and within agencies.

While all four of the agencies have taken steps to foster data sharing, they do not routinely monitor whether researchers make data available from all

climate change research programs and have not fully overcome key obstacles and disincentives to data sharing. For example, one key obstacle is the limits in the data infrastructure to preserve particular kinds of climate change data. Data archives for certain kinds of data in some disciplines, such as ecology, do not exist, which places a greater burden on the individual researcher to maintain and preserve data. Preparing data for future use is also a laborious and time-consuming task that can serve as a disincentive to sharing data. Furthermore, multiple data management officials said that data preparation does not result in the same benefits, such as career advancement, as publishing results in journals can. Officials also noted that researchers are expected to make underlying data available and to publish results in journals, but traditionally the scientific community has mainly rewarded publication. Consequently, researchers who have to compete for funding are more likely to focus on publishing research results than preserving underlying data for future use, thereby putting the data at risk of being lost or inaccessible to other researchers.

We are making recommendations to the federal climate change research agencies to improve their ability to ensure that federally funded research data are made available to other researchers. Specifically, we recommend that the Secretary of Commerce and the NOAA Administrator clearly inform researchers in writing of NOAA's data-sharing expectations. We also recommend that DOE, NASA, NOAA, and NSF consider steps for maximizing opportunities to encourage researchers to make data available to other researchers, including evaluating data-sharing plans when considering grant proposals. Finally, we recommend that the agencies evaluate whether additional strategies to facilitate permanent archiving of relevant data are warranted. In commenting on a draft of this report, the four agencies generally agreed with our findings and recommendations. Some of the agencies provided technical clarifications, which we have incorporated in this report as appropriate.

Background

The federal government has funded climate change programs for over 20 years, and the budget for climate change research and development—approximately \$5.9 billion in fiscal year 2006—supports a wide range of programs. As in the past, nearly half of the fiscal year 2006 federal climate change budget funded technology programs that focus on responses to climate change, such as developing and deploying technologies to reduce greenhouse gas emissions or increase energy efficiency. Less than one-quarter goes toward tax provisions related to climate change. These provisions encourage emissions reductions through, for example, tax incentives to encourage the use of renewable energy. A fraction of the

budget also contributes to international assistance programs that seek to help developing countries address climate change by, for example, improving energy efficiency technology. Finally, the estimated \$1.7 billion spent on science research programs accounts for roughly one-quarter of the total budget for climate change programs and are the focus of this report.

Federal climate change science programs seek to monitor, understand, and predict climate change through both agency-led and external research activities. In particular, the science programs seek to advance the state of knowledge on (1) natural climate conditions and variability; (2) forces that influence climate; (3) climate responses; (4) the potential impacts of climate change on the environment, population, and the economy; and (5) ways to apply this knowledge to decision making. A total of 13 federal departments and agencies support climate change research activities, though 4 of these departments and agencies—DOE, NASA, NOAA, and NSF—received about 90 percent of climate change science funding in fiscal years 2005 and 2006. NASA accounts for the greatest portion of the climate change science budget, about 61 percent, followed by NSF (12 percent), NOAA (9 percent), and DOE (8 percent).⁵ Agencies may also contribute funds from nonclimate specific accounts to the infrastructure supporting climate change research, such as sophisticated instruments and equipment. In particular, the climate change research budgets do not reflect all of the funds that NASA and NOAA contribute to satellite systems and sensors used to collect data. Also, DOE, NASA, and NOAA each have laboratories that perform climate change research. Unlike the mission-based agencies, NSF is a funding agency supporting all fields of fundamental science and engineering. NSF provides about 20 percent of all federally supported basic research conducted at U.S. colleges and universities and generally funds this work through limited-term grants issued to institutions supporting individuals and small groups of researchers.

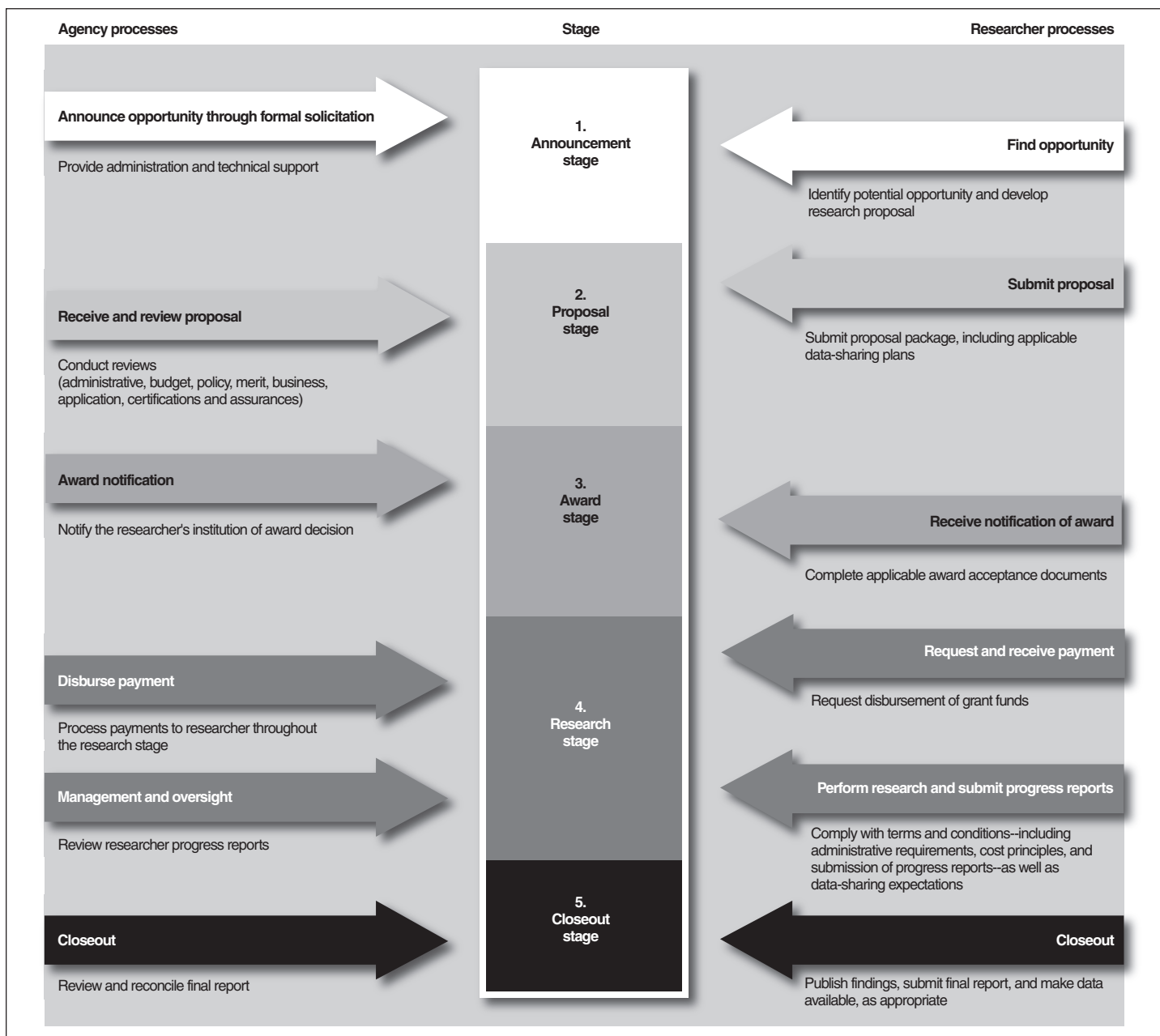
⁵These figures are based on the estimated total U.S. Climate Change Science Program budget—which includes both scientific research and NASA space-based observations—for fiscal year 2006. Estimates for spending in fiscal year 2007 were not available as of August 2007. See Climate Change Science Program and the Subcommittee on Global Change Research, *Our Changing Planet: The U.S. Climate Change Science Program for fiscal year 2007* (Washington, D.C., 2007).

All four agencies support external climate change research, primarily through grants.⁶ The grant review process typically begins when a researcher, or group of researchers, responds to an agency's formal solicitation with a written research project proposal. Such proposals generally summarize how the researcher would use grant funds to respond to the agency's solicitation, including how the researcher would perform the work as well as the budget and timeline for doing so. Some proposals may also describe plans to collect and manage data.

The agencies assess many proposals on a competitive basis. Usually a program manager who oversees many research grants assumes responsibility for the scientific, technical, and programmatic review of the proposals submitted in response to the solicitation. In addition to the intellectual and scientific merit, as well as the potential broader impact of the proposal, agencies may use criteria such as the past performance of the researcher, as well as the budget and priorities for the agency's program, when determining whether to fund the proposal. Also, agencies request written reviews or independent panels of the researcher's peers to assess the scientific merit of proposals in some cases. The program manager then recommends to other agency officials which proposals the agency should fund. The agency then compiles notification letters that formally offer the grant to the researcher's institution and outline the terms and conditions in the grant agreement, a legal instrument describing the relationship between the agency and the recipient (see fig. 1 for a summary of this process).

⁶For purposes of this report, grant refers to cooperative agreements as well as awards used by DOE to fund external researchers at National Laboratories.

Figure 1: General Research Grant Process



Source: GAO.

After agreeing to the terms and conditions of the grant, the researcher begins the work and submits periodic progress reports to the agency. The researcher's primary point of contact with the agency is the agency program manager who oversees the award. When the results of their investigation are ready, researchers usually attempt to publish their findings and conclusions in peer-reviewed journals. The publication of research results in journals can advance the state of science and benefit the researcher through, for example, career advancement.

According to a senior official at the journal *Science*, nearly all researchers seek to share their results and conclusions through journal articles, but as the National Academies and agency officials have acknowledged, the mechanisms of making the data underlying their results available to others can vary greatly and involve many different stakeholders. Accordingly, the expectations for data sharing can vary by research type. In the past, researchers generally kept the data they collected or generated under a grant award in their possession and made them available to other researchers upon request. The development of sophisticated tools and use of the Internet as a means to disseminate information has greatly expanded data-sharing opportunities. Researchers can submit some types of climate change data to federal archives that preserve electronic data online. Some of these archives are managed by federal programs separate from those funding climate change research. However, archival infrastructure does not exist for all kinds of climate change data. Indeed, the National Science and Technology Council has established an Interagency Working Group on Digital Data to develop and promote a "strategic plan...to ensure reliable preservation and effective access to digital data" derived from federally funded research. To ensure access to relevant data, some journals have developed online databases to store data that support the articles they publish. Researchers may also make data available by posting them on personal or institutional Web sites and, with physical samples, by housing the materials in facilities such as the National Ice Core Laboratory. When no archive or other mechanisms for making data available exists, researchers may store data in their own files and make them available to others upon request.

Determining what data to make available from past research activities can pose a challenge because data are not always static or discrete. A National Academies panel on Science, Technology, and Law described data as information that moves through many levels, ranging from raw data to final data, during the research process. Before they can be made available, researchers validate and perform quality assurance measures on the data by, for example, deleting outliers or coding the data for use in software

applications. Distinguishing between raw, processed, and final data is often a subjective determination and requires scientific and technical expertise.

Despite these differing expectations about how and what data to make available, the scientific community has long promoted data sharing.⁷ In particular, the National Academies have studied and promoted data sharing through a series of committees, symposia, and studies. Federal research agencies that fund data collection, professional scientific associations such as the American Geophysical Union and the American Meteorological Society, and academic journals such as *Science* and *Nature* have also produced a series of statements and policies on data sharing. Other information science scholars have published studies of data-sharing policies and practices as well. Though some of the work produced by this community has focused on data sharing within particular disciplines, such as the earth and life sciences, it believes that research data should generally be shared and available to all researchers.

The Scientific Community Has Identified Several Key Issues That Policies Should Address to Facilitate Data Sharing

The National Academies, professional scientific associations, and other members of the scientific community, have identified key issues that data-sharing policies should address, including what, how, and when data are to be shared, as well as the cost of making data available. The way in which specific policies address key data-sharing issues, however, may vary depending on the type of research. Policies must take into account their applicability to specific research projects, relevant legal and regulatory restrictions, the existence of appropriate archives, and the characteristics of particular research fields.

⁷For the purposes of this report, the scientific community refers to the general body of scientists and its institutions as represented by the National Academies and professional scientific associations. While no single body can be said to speak for all of science, the National Academies and other scientific associations, such as those listed in appendix I, often act as surrogates when the opinions of the scientific community, or particular disciplines within science, need to be ascertained. We also supplemented our analysis of the reports and statements of these organizations with interviews of officials, at a variety of entities, with knowledge of data-sharing issues. Furthermore, whenever we attribute statements to the scientific community at large, we mean that the National Academies and at least two of the scientific associations listed above support those statements.

Data-Sharing Policies Should Address What Data Are to Be Shared

The scientific community generally believes that data-sharing policies should address what data are to be shared and that, at a minimum, the information necessary to support researchers' major published results should be made available to other researchers. The National Academies have recommended in, for example, a 1997 report from the National Research Council that federal science agencies adopt, as a fundamental goal, the full and open exchange of scientific data derived from federally funded research. Various scientific associations, such as the American Association for the Advancement of Science and the American Geophysical Union, have also identified the open availability of data as an issue that data-sharing policies should address and support. These organizations have supported open access to research data because of the many benefits of sharing data. Open access to data, according to these organizations, maximizes the societal benefits of the scientific endeavor. Moreover, when data are widely available, the information can be used to provide a direct check on reported results or advance future research in a field of study. According to officials with whom we spoke at archives and the National Academies, data sharing can be particularly important in the field of climate change research, because accessing data from a variety of sources is crucial to understanding the multivariate nature of the earth's climate. Officials also emphasized that information made available to the wider research community should include both the raw data or physical samples resulting from the research as well as the metadata—i.e., information needed to understand the content, quality, and condition of the data—because both the raw data and metadata are essential for other researchers to make practical use of shared information. In addition, NOAA stated that in all cases sufficient metadata, such as data set descriptions, should be provided so the data can be found and their suitability for use determined.

Though the full and open exchange of data is supported as an overall goal, the scientific community acknowledges that there are certain legally binding limitations to the goal of openness. In particular, there are statutory and other legal limits on data sharing designed to protect intellectual property, privacy, and national security. Protecting the privacy of human subjects and national security have been acknowledged as legitimate limitations to the full and open exchange of scientific data by the National Academies. More recently, according to the National Academies, protecting intellectual property has created new restrictions on data sharing. National laws and international agreements in the area of intellectual property rights, privacy, and national security may directly affect data access and sharing policies. Scientific associations have also recognized these constraints to data sharing, while also noting that the

majority of data collected with public funds are not affected by these restrictions.

Practical limitations, such as a lack of appropriate archives for storing data, can also affect how policies address the goal of openness. If no archive exists, then researchers may not be able to make their data available on a long-term, low-cost basis. Moreover, policies applicable to research that produces modeling or experimental research data may not require all results to be shared. While modeling activities generate large volumes of data, only a portion has an appropriate archive or is useful to the wider research community; therefore, some model data are generally retained by the researcher and made available to others upon request. Furthermore, according to a senior official at the National Academies, experimental research data are created as a result of a specific process or analysis and can often be recreated, so sharing the actual data and materials is not as important as sharing information about research methodologies. Raw observational data, on the other hand, are unique and if not made accessible to other researchers may be lost forever. Thus, the scope and methods for sharing data generally depend on the type of research that was conducted. Table 2 in appendix III provides examples of how data-sharing expectations can differ by research project type.

Data-Sharing Policies Should Address How Data Are to Be Shared

The scientific community generally believes that data-sharing policies should address how data are to be made available and that, when the appropriate infrastructure exists, data acquired in federally funded research should be made accessible through unrestricted archives. Many existing data-sharing policies and guidelines encourage researchers to place their data in public archives. The National Academies have recommended in multiple reports that these data be made readily accessible—ideally via the Internet—through repositories that are supported by a community of researchers and in general use. Scientific associations have also acknowledged that data-sharing policies should be guided by the goal of making data available for the long-term via archives.

The lack of appropriate infrastructure for the sharing and preservation of certain kinds of data may affect specific data-sharing policies, particularly for federal agency research programs. Some scientific disciplines, such as ecology and hydrology, do not have the infrastructure to facilitate data sharing. Furthermore, research performed by individual investigators or small research groups operating outside large research programs may not have appropriate data archives. Without such archival infrastructure,

researchers working in these fields or research programs may not be able to easily share their data with others.

Data-Sharing Policies Should Address When Data Are to Be Shared

The scientific community believes that data-sharing policies should address when data should be made available and that data should generally be made available immediately or after a limited proprietary period that allows researchers to complete their initial analysis and publish their results. In an early report on data sharing by the National Research Council, the National Academies recommended that research data be made available by the time the initial major results are published, except in compelling circumstances. Further, the report maintained that data relevant to public policy should be shared as quickly and widely as possible. Various scientific associations also support the goal of making data available to the public as early as possible.

While immediate open access to data is desirable, the premature disclosure of research data may disrupt the processes of analysis, interpretation, and peer review that normally precede such public disclosure, according to the American Association for the Advancement of Science. Accordingly, a federal agency scientist told us that the research community recognizes the need for researchers to perform quality checks on data and publish their results before releasing the data to other researchers. Indeed, a limited proprietary period for principal researchers is a common principle in the research community. However, the duration of such a period may be determined by the type of research. In particular, the length of the proprietary period in which a researcher, or a group of researchers, has exclusive access to data may vary by research project or discipline. Some research projects, such as those gathering observational data from satellites, are often expected to make their data available immediately after the standard period of calibration of equipment and validation of observations. Alternately, other projects, such as those involving the collection of physical samples, cannot make their data available immediately due to logistical constraints. Ice core samples, for instance, cannot be made widely available due to the fact that they are difficult to transport and must be stored in a particular, central location. Moreover, certain projects involve a substantial investment of time and resources by the researchers, and it is generally agreed that such researchers are entitled to a period of exclusive use of the data they have collected. Such projects are, in some cases, granted a proprietary period of up to 2 years to allow researchers to develop publishable results and prepare the data for sharing. Still other projects, including those involving multiple researchers, make their data available among original researchers

immediately and to other researchers after a limited proprietary period. According to a 1997 National Academies report on data access issues, the maximum length of any proprietary period should be established by particular scientific communities. Moreover, scientific associations have acknowledged the need for differing proprietary periods and called for federal agencies to tailor their data-sharing policies and expectations to specific research projects when necessary.

Data-Sharing Policies Should Address the Cost of Making Data Available

The scientific community generally agrees that data should be made available at no cost or for no more than the marginal cost of reproduction and distribution. Moreover, the process of sharing data should seek to minimize the burden to researchers of making data available. The National Academies and various scientific associations recommend the full and open exchange of data, including making federally funded research available for no more than the cost of reproduction and distribution. According to the American Geophysical Union, this goal is designed to balance the costs associated with sharing data with the desire to make data easily accessible, so as to not impose significant burdens on original or subsequent researchers. Government agencies have often charged fees for access to data in order to recover the costs of generating or reproducing the data. However, with the reduced costs of capturing and storing digital data, agencies are now often able to provide data for no cost on the Web. Indeed, in a 2003 report, the National Academies recommended making federally funded data available for research purposes at no cost when possible. Nevertheless, since the cost of sharing data will likely depend on the type and format of the data, archived data not available digitally—such as physical samples—may involve higher costs for original or subsequent researchers.

Climate Change Research Agencies Rely on Various Policies and Practices to Encourage Researchers to Make Data Available

All four agencies said that they adhere to governmentwide data-sharing guidelines and, to varying degrees, have their own agency, program, and project-specific data-sharing policies. The manner in which these policies address key data-sharing issues like openness, timing, and cost vary among and within agencies based on the needs of specific research programs. Agencies also facilitate data sharing through unwritten practices, such as providing incentives for data sharing through the grants process, maintaining personal contact with researchers, and encouraging researchers to archive data.

All Four Agencies Said They Adhere to the Governmentwide Data-Sharing Policy

While federal statutes do not clearly specify data-sharing requirements for external climate change researchers using federal funds,⁸ some program managers at each agency reported in our survey that they had incorporated requirements into particular grants. The agencies, however, have relied primarily on a number of policies and practices to encourage data sharing among external researchers. At the broadest level, the agencies recognize an interagency policy on climate change data, which represents a governmentwide commitment to make climate change data available to other researchers.⁹

Specifically, the *Data Management for Global Change Research Policy Statements*, an interagency policy under the Climate Change Science Program (CCSP), provides guidance to the agencies on how to ensure that researchers make federally funded climate change data available to researchers.¹⁰ A related interagency research group that predates the CCSP—the Global Change Research Program—developed the policy in response to concerns that inadequate attention was given to maintaining climate change data. The Global Change Research Program observed that “proper preparation, validation, description, and care of data sets is critical to their use by the widest possible scientific community.” The CCSP has encouraged those agencies funding climate change research to incorporate the guidelines listed in this voluntary policy into their data-sharing policies and practices. Senior officials at DOE, NASA, NOAA, and NSF told us that their data-sharing policies and practices adhere to the principles of the guidelines.

The interagency policy addresses the key issues of data sharing, such as openness and accessibility. For example, the policy’s overarching objective calls for the “full and open sharing of the full suite of global data sets” by all climate change researchers. The policy further specifies what

⁸OMB A-130 is a governmentwide policy calling for “the open and efficient exchange of scientific and technical information.” See OMB A-130, *Management of Federal Information Resources* 7(k) (Feb. 8, 1996). OMB A-130 focuses on the information activities of all agencies of the executive branch of the federal government, but does not clearly specify responsibility of federally funded researchers or that of the government to foster data sharing under grants.

⁹As stated earlier, we define data to include factual information or physical samples that are collected and recorded as a result of scientific observation, experiment, analysis, or similar methods of research. The output of models can also be considered data.

¹⁰U.S. Climate Change Science Program and the Subcommittee on Global Research, *Strategic Plan for the U.S. Climate Change Science Program* (Washington, D.C., 2003).

counts as data and broadly defines them as the information “resulting from observations, the application of algorithms to produce new data, and from the data output of models.” The policy states that metadata should be made available to allow researchers to assess the quality of data. In addition to encouraging open data sharing among all climate change researchers, the policy addresses accessibility by recommending the long-term preservation of data in archives. The policy states that agencies funding research should develop procedures and criteria for obtaining, maintaining, and purging data in the archives. See appendix IV for more detailed information on how the policy addresses key data-sharing issues.

Data-Sharing Policies Vary among and within Agencies

Our review, which included a survey of 64 program managers at the four major climate change research agencies, identified 23 different policies¹¹—accounting for about 80 percent of the agencies’ climate change research programs—that encourage researchers to make data available.¹² Although data sharing is generally regarded as a standard practice among colleagues, the mechanics of data sharing—such as what data to preserve and when—involve some professional judgment. To guide researchers, the four major climate change research agencies have policies that document these mechanics. Agencies’ written policies emphasize their commitment to data sharing and standardize expectations for data sharing. Overall, the policies range from broad statements calling for open and timely access to data to more detailed policies that define the mechanisms and timelines for making data accessible. NASA and NSF have data-sharing policies documented at the agency level that address openness and timing and apply to all topics of research; all four agencies have various program and project-specific data-sharing policies.

While the policies generally underscore the importance of making data openly available at minimal cost, we found that they vary among and within agencies because they are often tailored to the needs of different research programs or projects within the same agency. Accordingly, these policies address in different ways the key issues discussed in the previous section of this report. For example, variations in archiving resources and

¹¹For purposes of this report, policy refers to written, nonbinding agency directives and guidance intended to inform agency managers, staff, and researchers.

¹²Data-sharing policies were identified for 40 of the 51 different programs. The number of discrete policies differs from the number of programs because some policies applied to more than one program, and some programs had more than one policy.

the extent of quality assurance required—such as validation and calibration—influence a policy’s recommended time frames for data sharing. See appendix II for a complete list of the data-sharing policies.

DOE’s climate change research programs have established written policies that encourage researchers to make data available within certain time frames and according to specific standards. One of DOE’s programs that funds the collection and analysis of measurements from instruments through a network of researchers issued a policy that encourages researchers to make data available quickly. This particular DOE policy distinguishes between data that have been quality assured and preliminary data, which have not been validated, and affords researchers some time to work on the data before finalizing data submissions to an archive. However, DOE expects that even the preliminary data will be made available almost immediately. The policy calls for “near real-time” sharing of the data among the researchers participating on the team and for making the data available to the research community within days to allow for routine processing and electronic archiving.

NASA’s agencywide policy briefly states that researchers should make data available at the earliest possible time, whereas its earth science program provides greater detail about what data to share and when to do so. For example, NASA’s earth science program policy states that researchers do not have a period of exclusive access to the satellite data, which are made available in the agency’s data archive system as soon as they are calibrated and validated. According to senior NASA officials, the program formerly granted researchers a 2-year period of exclusive use of the data but determined that the wider benefits of making data available to all outweighed the benefit of temporary restrictions. One NASA official, however, noted that there is a trade-off, as the lack of an exclusive use period limits opportunities for researchers to analyze the data and make them more user friendly.

The NSF agencywide policy states that researchers are “expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered.”¹³ In order to address the needs of specific research programs, program-level policies often provide researchers more detailed guidance about how to carry out the

¹³National Science Foundation, *NSF Grant Policy Manual*, (Arlington, VA, 2005).

agencywide data-sharing policy. This agencywide policy establishes a general expectation that data are to be shared with other researchers. The data-sharing policy for the oceans program—one of NSF’s programs funding particular climate change research—identifies particular archives for researcher use, such as one that preserves sediment samples from the ocean floor. Further, the agencywide policy states that data are to be shared “within a reasonable time” and the oceans program policy states that data should be shared as soon as possible but no later than 2 years after collection.

The Climate Observation Program is NOAA’s only climate change research program that has issued a written data-sharing policy. Similar to the DOE example given above, NOAA’s Climate Observation Program policy states that researchers should make data available near-real-time with associated metadata and free of charge to others. The policy further notes that the data should be made available quickly enough to “be of value to operational forecast centers, international research programs, and major scientific assessments.”¹⁴

We found that data-sharing policies vary in part because the type of data generated differs by program. An official with the American Association for the Advancement of Science observed that variations in data-sharing policies by data type reflect the differences in the ways data are collected and accessed. According to one survey respondent, data generated instantaneously—such as meteorological data from instrument measurements—may not require as much preparation or quality assurance as other forms of data, like physical samples, that may require extensive analysis and interpretation. For example, according to a DOE official, large atmospheric science data sets generated by DOE-funded researchers require supercomputers for analysis and therefore require more time and processing before the researcher can transfer the data to someone else.

Furthermore, we found that agencies also have project-specific data-sharing policies. The AmeriFlux program—a network of climate change researchers funded by multiple agencies, including DOE, NASA, NOAA, and NSF—provides another example of how the agencies tailor data-

¹⁴National Oceanic and Atmospheric Administration, Climate Observation Program, *Data Policy*, <http://www.oco.noaa.gov/>.

sharing policies to the needs of particular projects.¹⁵ The AmeriFlux program requires participants to submit data to a designated archive within 1 year of collection and specifies the preferred format for data submission. According to program officials, the 1-year limit allows researchers to spend time preparing and documenting the data in adherence to the standards specified by the archive that are intended to facilitate access and use by other researchers.

We also found that large, collaborative projects like AmeriFlux usually establish data-sharing policies for the participants. These projects typically involve multiple funding agencies and researchers based in different locations, some even in different countries. Similar to the program-level policies, the project-specific policies tailor the agencies' expectations for data sharing to the data management needs of the project. For example, the NSF- and NOAA-funded U.S. Global Ocean and Ecosystems Dynamics project, which involves collaboration among physicists, biologists, chemists, meteorologists, and resource managers, established a policy to guide participants' data sharing. The policy describes data sharing as an iterative process and instructs researchers to work with the data managers to assess what data would be most important to share. The policy also encourages researchers to submit data to an archive and to include metadata to facilitate their use by others.

Agencies Also Rely on Unwritten Data-Sharing Practices

While many program managers described written policies for data sharing as essential to advancing the state of climate change science, they also identified unwritten practices that the agencies use to encourage and facilitate data sharing. The flexibility of practices allows program managers to tailor data sharing to the needs of a specific project. These data-sharing practices include using the grants process to provide incentives for data sharing, maintaining personal contact with the researchers, and archiving data. Our review shows that use of practices varies among and within the agencies.

Grants Process and Personal Contact Are Used to Encourage Data Sharing

The agencies use the grants review process to provide incentives for data sharing, thereby encouraging researchers to make data available. Indeed, some program managers use the evaluation of grant proposals as an

¹⁵While multiple agencies fund the research and analysis conducted within the AmeriFlux Network, an official at the AmeriFlux archive reported that DOE provides nearly all of the funding to manage the data at the archive, which is based at the Carbon Dioxide Information Analysis Center at Oak Ridge National Laboratory.

opportunity to encourage researchers to identify and plan in advance for data management needs—such as how they will preserve and make data available. We found that NSF expects researchers applying for grants to present, as appropriate, a clear description of “plans for preservation, documentation, and sharing of data, samples, physical collections, curriculum materials, and other related research and education products.”¹⁶ However, the general grant guidance materials for researchers applying for DOE, NASA, and NOAA climate change grants do not explicitly instruct them to include data-sharing plans in their proposals. Nevertheless, some program managers encourage researchers to do so in practice.

DOE and NASA officials told us that program managers might encourage researchers to include data-sharing plans on an ad-hoc basis. An example of this practice, according to DOE, is to request the data-sharing plan in the solicitation notice for a particular award. DOE and NASA officials could not confirm the frequency of this practice, however. The extent to which program managers can use data-sharing plans as a criterion for grant award decisions appears limited because most of the climate change research programs do not explicitly require them.

Funding decisions made throughout the grant process are also used by agencies to hold researchers accountable to data-sharing expectations. Most of the program managers we surveyed reported that they consider researchers’ past data-sharing practices when deciding whether to fund research proposals. Once the agency has awarded the grant, program managers may use the staggered installments of grant funds as another incentive to encourage researchers to make data available to others. Two program managers reported in our survey that they withhold funding installments if researchers have not made data available.

The extent to which federal climate change research agencies use various aspects of the grant review process to encourage data sharing varies, depending on the initiative of the program manager, in part because there are no requirements for them to do so. For example, an NSF official stated that the consideration of past data-sharing activities is not a discrete factor that the agencies require program managers to use in making award decisions.

¹⁶National Science Foundation, *NSF Grant Proposal Guide*, (Arlington, VA, 2004).

Agencies Also Rely on Data Archiving to Foster Data Sharing

Moreover, the agency officials told us that they have limited information about whether researchers make data available. Some program managers said that they attempt to determine whether researchers are making data available by reviewing progress reports—required written updates submitted by researchers. Progress reports inform program managers of the status of research throughout the duration of the grant, and a final report documents completion of the research. The final progress report typically states whether the researcher has published the results in a journal. However, progress reports do not necessarily provide program managers enough information to assess the availability of the data.

Many program managers reported in our survey that they maintain personal contact with researchers to ensure that they make data available to other researchers.¹⁷ Such personal contact can serve to remind researchers of the importance of making data available and help them address any difficulties in doing so. One agency official noted that program managers often rely on personal contact to encourage researchers to make data available to others. Some program managers also reported that they would authorize additional funds to help ensure that data sharing occurs.¹⁸

Data archiving is one of the primary data-sharing practices used by the federal climate change research agencies, according to our survey and interviews with agency officials and data management experts. Archiving refers to the long-term storage of data, most often in digital form, but there are also some repositories that can hold physical samples. A majority of program managers surveyed identified archiving as one way that researchers make data available. Climate change research directors at all four agencies also said that the agencies encourage researchers to use archives to make data available.

In addition to encouraging researchers to archive data, the mission-based agencies—DOE, NASA, and NOAA—support archiving practices by operating permanent data archives that store data, such as satellite images and measurements of indicators in the atmosphere, land, and oceans used

¹⁷Of the 55 program managers who indicated that their program takes steps to ensure researchers make data available, 49 said they do so by maintaining personal contact.

¹⁸Thirty-seven of the 55 program managers indicating that their program takes steps to ensure researchers make data available said they would do so by authorizing additional funds.

to understand climate change. According to senior officials, these agencies have made notable financial investments in these data management and preservation services.¹⁹ For example, according to a DOE official, the agency contributes about \$2.7 million annually to the Carbon Dioxide Information Analysis Center, which preserves data from researchers collaborating in the AmeriFlux Network and other climate change programs.

The archives, typically managed by programs separate from those sponsoring climate change research, provide a number of services to facilitate data sharing. For example, staffs operating the agencies' data centers do not only permanently archive data in electronic databases, but they also perform quality control measures to standardize and make data usable to a wide audience, develop data products to facilitate additional analysis, and help other researchers navigate the database to find relevant information.

While part of the agencies' investment in archives support data managers, generally neither the agency nor the data managers actively solicit data from the researchers. One exception identified by a survey respondent is NOAA's Climate Prediction Program for the Americas, which employs a manager who collects data from researchers, performs quality control measures on the data, and make data available on a Web site. A second exception is the outreach conducted by staff at an NSF-funded archive, the National Center for Atmospheric Research (NCAR) Coupling, Energetics, and Dynamic Atmospheric Regions. Another survey respondent said that NCAR data managers collect instrument measurements from researchers funded by one of NSF's climate change programs. As part of its efforts in maintaining the archive, NCAR sends reminders to researchers to submit data, and as necessary, notifies NSF program managers of researchers who have not submitted data. NSF follows up with the researchers to ensure data are submitted to the archive.

However, not all data can be digitally archived because, for example, data such as physical samples may not be in a form amenable to this type of storage or because archives do not exist for the data from some types of climate change research. Agencies recognize this challenge and have relied on other practices to encourage data sharing. For example, most

¹⁹NSF officials reported that they have provided some financial support for data archives but that NSF does not fund archives for climate change data on a permanent basis.

program managers (59 of 64) reported that researchers publish research results in journals that indicate where to find the underlying data.

All Four Agencies Have Taken Steps to Foster Data Sharing but Have Not Fully Overcome Key Obstacles

While the agencies have taken steps to foster data sharing, the effectiveness of their requirements, policies, and practices is unclear because the agencies do not routinely monitor researchers' data-sharing activities. As a result, the agencies lack information to assess the extent to which researchers are making federally funded climate change data available. In addition, we found that the agencies have not fully overcome key obstacles and disincentives to data sharing that could limit the availability of data.

Agencies Do Not Routinely Monitor Whether Researchers Make Data Available

While senior officials at all four agencies believe that researchers share the data derived from federally funded research projects, the effectiveness of their data-sharing requirements, policies, and practices is unclear because the agencies do not routinely monitor whether researchers make data available from all climate change research programs. Instead of proactively overseeing data sharing, the agencies rely on self-policing within the research community. That is, they assume that researchers will adhere to the norms of data sharing and expect members of the research community to notify them when researchers do not make data available. According to our survey, roughly one-third of the 64 program managers have, within the past 10 years, been notified that an award recipient did not make data available. Nearly all of the program managers said they responded to the reported problem, and many believed it was resolved.

Although researchers can contact the agency if other researchers withhold data, this is not an effective way to resolve situations involving incomplete or missing data. Several data managers told us that documentation about the data—such as conditions under which it was gathered—is crucial because important details about the data are likely to be forgotten as the researcher moves on to new projects. Furthermore, at some point, it may become too late for federal agencies to encourage data sharing because by the time one requests access to certain data—possibly years after the initial data collection—the original researcher may have lost the data or failed to record important metadata. Therefore, we believe that agencies' reliance on self-policing by the research community does not provide adequate assurances that researchers will fulfill the data-sharing expectations set forth in the agencies' policies.

Senior agency officials at all four agencies told us that it is impractical for program managers to verify data sharing because they oversee many researchers and must focus on higher priority tasks. Moreover, several of these officials believe that current self-policing is effective because of the collaborative nature of climate change research. The agencies fund many large climate change research projects that involve multiple researchers who depend on one another to share data in a timely manner. The researchers participating in such projects typically submit data to an archive and also hold one another accountable. For example, senior DOE and NASA officials reported that they convene science team meetings wherein they coordinate activities and receive updates from the funding agency. According to a senior DOE official, meeting participants address data-sharing issues at these meetings. He noted that the meetings provide a particularly effective forum for researchers to call attention to those who have not made data readily available. However, there appears to be greater accountability among researchers collaborating with one another on similar projects than among researchers who work on individual projects.

Researchers seeking data that have not been made widely available, such as through an archive, generally need to contact the original researcher(s) to request data. While most of the program managers we surveyed indicated that there are several incentives for researchers to make data available—such as maintaining informal relationships with other researchers, obtaining recognition in the scientific community for the work, or the potential for future collaboration—there is no guarantee that the original researcher will have the complete data readily available to comply with another researcher’s request for data. Furthermore, researchers face a number of practical obstacles that may limit their ability to document and preserve data.

The Agencies Have Not Fully Overcome Key Obstacles and Disincentives to Data Sharing

Despite the various incentives for researchers to make data available to others, there are several obstacles and disincentives to data sharing that the four agencies have not fully overcome. For example, one key obstacle is the limits in the data infrastructure, such as the lack of archives capable of preserving certain kinds of climate change data being generated by federally funded research. Data centers funded by DOE, NASA, and NOAA currently archive digital data; some data centers preserve physical samples such as ice cores or ocean sediments. Archives currently in operation store data from some areas of climate change, including oceans and atmospheric sciences; but according to officials at NSF, the National Research Council, and several scientific societies, permanent repositories

are not available for other fields within climate change, such as certain kinds of ecological and earth sciences data.

According to several data management stakeholders, the options available to preserve data, such as electronic archives, are limited for climate change data developed through the use of computer models. While there are some archives that store data from climate change models, such as the DOE-funded Program for Climate Model Diagnosis and Intercomparison, these stakeholders told us that permanent model data archives are generally lacking. Furthermore, the limits in data infrastructure for climate change data create a greater burden for federally funded researchers to maintain and preserve data themselves. The National Academies have raised concerns about the long-term availability of federally funded data and observed in one report that “data sets that commonly are gathered at great expense and effort are not broadly available and ultimately may be lost, squandering valuable scientific resources.”²⁰ The report concluded that funding agencies should be responsible for making the data available to others.

The four agencies also have yet to effectively address key disincentives to data sharing on the part of researchers. For example, the time and labor required to prepare data are significant disincentives to making data available for other researchers. One program manager commenting on the practical obstacles to data sharing noted that while most researchers are willing to share data, they “resist the large additional costs of time or money to meet requirements.” Making data available often involves laborious and time-intensive tasks to adequately document the data and to perform quality assurance checks, such as correcting errors, to make them usable for other researchers. For example, the National Academies have recognized an administrative and cost burden that largely falls on the researcher to prepare data for others’ use. Researchers may also need to summarize the data processing history, develop a codebook, and write instructions on how to use the data files.

Moreover, researchers must weigh the trade-offs in costs and benefits, according to one program manager, such as the limits of the program budget and whether responding to detailed requests would impede

²⁰National Research Council, *Preserving Scientific Data on Our Physical Universe: A New Strategy for Archiving the Nation’s Scientific Information Resources*, (Washington, D.C.: National Academies Press, 1995), p. 3.

progress on additional research. Some of the directors of the climate change research programs raised similar concerns about research priorities, in light of resource constraints. A senior DOE official noted that while the agencies can encourage researchers to make data available, funding priorities do not typically favor the time-consuming tasks involved in making data available. The official clarified that when faced with budget constraints, agencies tend to target limited funds to new visible research at the expense of data archiving. The National Academies also recognized the bias toward new research projects and found that among all scientific disciplines, most agencies make data management and preservation a low priority, even when the benefits of making data available from old projects exceed those realized from new projects.

Furthermore, multiple data management officials pointed out that researchers do not receive the same benefits, such as career advancement or peer recognition, for preparing data as they do from publishing research results in journals. These officials stated that funding agencies and the scientific community expect researchers to both publish their results and make underlying data available, but researchers have traditionally been rewarded mainly for publication. According to a National Academies report on data access, “society fellowship and award committees generally do not place much value on the contributions their applicants may make to the infrastructure of science in the form of data compilation, organization, and evaluation work.”²¹ As a result, researchers who have to compete for funding are more likely to focus on publishing research results than preserving underlying data for future use, thereby putting the data at risk of being lost or inaccessible to other researchers.

Our survey identified several additional disincentives that may deter data sharing, at least temporarily, including requests for more time to analyze data and concerns about intellectual property.

Conclusions

Government agencies articulate expectations for recipients of federal grants about important functions such as data sharing through written policies. Written policies both show that the agency views data sharing as a priority and facilitate researchers’ understanding of specific expectations about the mechanics of data sharing, which typically involve some professional judgment to determine, for example, what data to preserve,

²¹National Research Council, *Bits of Power* (1997), p. 61.

how to make it widely available, and the time frame for doing so. One particular collaborative program funded by DOE, NASA, NOAA and NSF, known as AmeriFlux, has written requirements designating the archive where researchers must submit their data as well as the time frame and preferred format for these submissions, all of which facilitate efficient data sharing by its participating researchers. This written policy helps ensure that all participating researchers understand the expectations and make data available in a way that advances the goals of the project. Similarly, written data-sharing policies exist under most federal climate change research programs at DOE, NASA, and NSF. Most of the research programs at NOAA, however, have not documented the agency's data-sharing expectations. Agencies such as NOAA that do not have a written policy at either the agency or program-level have fewer assurances about a mutual understanding of data-sharing expectations.

Federal agencies also use the grant review process to encourage data sharing by researchers. In some cases, the agency requires researchers to submit data-sharing plans in their grant proposals but the extent to which they use this as a criterion for grant award decisions appears limited. Once the agency has awarded a grant, program managers may use the staggered installments of grant funds as leverage to encourage researchers to make data available to others. Some program managers have effectively withheld funding installments when researchers do not make data available, while others review progress reports to determine whether researchers are making data available, taking action where they find instances of delay. In addition, during the grant review process, some officials informally consider researchers' past data-sharing practices in their evaluation, which conveys the importance of sharing research results among those involved in the research process. However, agencies have not institutionalized the use of the grants process to further data sharing and such efforts currently depend largely on the initiative of individual program managers who often oversee large grant portfolios.

The four research agencies we examined have policies and employ practices that encourage data sharing, which is ultimately the responsibility of the researcher. The agencies generally do not monitor and keep track of whether researchers make federally funded research data available. While the agencies believe that their data-sharing requirements, policies, and practices are effective, this is largely because they rarely receive reports suggesting otherwise. However, without data on actual data sharing by researchers, agencies cannot be sure their policies are working or determine whether changes in these policies are warranted. Measuring progress toward a goal of data sharing can allow

agencies to adjust their efforts over time to ensure that data are widely available to other researchers.

There are a variety of practical obstacles and disincentives to researchers sharing their data. Infrastructure is limited for storing data, such as that developed through computer models; and some fields of science, such as ecology, do not currently have archives in place that could maintain and preserve certain data. While developing and maintaining archives is an expensive undertaking, it is extremely important in areas of research related to climate change. Scholars from the National Academies and elsewhere have acknowledged the need to consider devoting additional research funds for the preservation of research data so that these valuable scientific resources, commonly gathered at great expense and effort, are broadly available to foster further research and analysis of long-term issues such as climate change.

Recommendations

To assist federal agencies sponsoring climate change research to better ensure the availability of data from federally funded research, we are making the following four recommendations.

To ensure that researchers receiving federal funds to conduct climate change research understand NOAA's expectations for data sharing, we are recommending that the Secretary of Commerce and the NOAA Administrator:

- Develop a set of written guidelines or use existing governmentwide guidelines, such as those endorsed by the Climate Change Science Program, to clearly inform researchers of NOAA's general expectations for data sharing.

To ensure that the agencies maximize opportunities to make data available in a manner useful to other researchers, we recommend that the Secretaries of Commerce and Energy, the NASA Administrator, the NOAA Administrator, and the NSF Director consider the following actions:

- Develop mechanisms for agencies to be systematically notified when data have been submitted to archives, so that agency officials have current information about the extent of data availability in order to adjust data-sharing policies over time to best meet the needs of researchers and the communities that use their data.

-
- Use the grant review process, where their program offices are not currently doing so, to facilitate further data sharing by (1) evaluating researchers' data-sharing plans as part of the grant review process and (2) using evidence of researchers' past data-sharing practices to make future award decisions. The use of such criteria in the grant review process should be clearly conveyed to researchers before they submit research proposals and after award decisions have been made.

To ensure that researchers make climate change data available to other researchers, we recommend that the Secretaries of Commerce and Energy, the NASA Administrator, the NOAA Administrator, and the NSF Director:

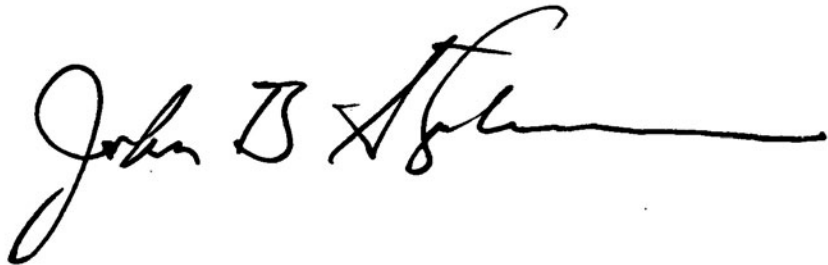
- Evaluate whether additional strategies are warranted to facilitate the permanent archiving of relevant data, which may include: leveraging existing resources; devoting a greater portion of data collection funds to archiving activities; or working with existing entities such as the National Science and Technology Council's Interagency Working Group on Digital Data, to develop additional data archives.

Agency Comments and Our Evaluation

We provided draft copies of this report to DOE, NASA, NOAA, and NSF. The four agencies generally agreed with our findings and recommendations. In addition, several agencies offered specific comments and technical clarifications, which we have incorporated in this report as appropriate. The written comments submitted by DOE, NASA, and NOAA are presented in appendixes V, VI, and VII; NSF provided technical clarifications orally. DOE commented on the importance of defining "data" and questioned whether we considered samples as data for purposes of this report. Our draft report included a definition of data that we have repeated, in an appropriate context, in an additional section of the final report. This broad definition, which includes research samples, allowed us to obtain a wide perspective on the variety of data-sharing requirements, policies, and practices. As we note in the report, each data-sharing policy may have different definitions of what data need to be shared. We agree that policies for physical samples will differ from those for electronic data, but we believe that each agency should make those determinations at the appropriate level.

As agreed with your offices, unless you publicly announce the contents of this report earlier, we plan no further distribution until 30 days from the report date. At that time, we will send copies to interested congressional committees and Members of Congress, the Secretaries of Commerce and Energy, the NASA Administrator, the NOAA Administrator, and the NSF Director. We also will make copies available to others upon request. In addition, the report will be available at no charge on the GAO Web site at <http://www.gao.gov>.

If you or your staff have questions about this report, please contact me at (202) 512-3841 or stephensonj@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made key contributions to this report are listed in appendix VIII.

A handwritten signature in black ink, reading "John B. Stephenson". The signature is fluid and cursive, with a long horizontal stroke at the end.

John B. Stephenson
Director, Natural Resources
and Environment

Appendix I: Objectives, Scope, and Methodology

Our objectives in this study were to determine (1) the key issues identified by the scientific community that data-sharing policies should address in order to facilitate the sharing of data from federally funded climate change research; (2) the requirements, policies, and practices to make data available to other researchers that exist under current federal climate change research awards from the four major federal climate change research agencies; and (3) the extent to which the major agencies effectively foster data-sharing. We defined the major federal climate change research agencies as those representing the bulk of federal climate change research spending. The Department of Energy (DOE), National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), and the National Science Foundation (NSF) represented nearly 90 percent of the U.S. climate change research budget in fiscal year 2006.

To address the first objective, we reviewed data sharing requirements and policies at federal agencies that were not identified by GAO as one of the four major federal climate change research agencies, such as the National Institutes of Health. We also reviewed the data-sharing policies at academic journals, including the *American Economic Review*, *Bulletin of the American Meteorological Society*, *Econometrica*, *Geophysical Research Letters*, *Global Biogeochemical Cycles*, *Journal of Applied Econometrics*, *Journal of Applied Meteorology and Climatology*, *Journal of Atmospheric Science*, *Journal of Climate*, *Journal of Geophysical Research*, *Journal of Physical Oceanography*, as well as the journals *Nature* and *Science*. We reviewed these particular journals because they either have an explicit data-sharing requirement or were identified in our survey of agency program managers as a leading publisher of climate change research. We also reviewed the data-sharing policies, statements, and reports of professional scientific societies, including the American Association for the Advancement of Science, American Geophysical Union, American Meteorological Society, Ecological Society of America, Geological Society of America, International Council for Science, and the U.S. National Academies. These associations were chosen because they either represented a broad cross-section of the scientific community or represent researchers in disciplines related to climate change research. Beyond an examination of the written policies and statements, we conducted interviews with officials at these organizations to gather additional information on data-sharing goals, practices, and issues. We also conducted a literature search to identify relevant studies of data-sharing policies, practices, and challenges. For the purposes of this report, the scientific community refers to the general body of scientists and its institutions as represented by the National Academies and professional

scientific associations. While no single body can be said to speak for all of science, the National Academies and other scientific associations such as those listed above often act as surrogates when the opinions of the scientific community, or particular disciplines within science, need to be ascertained. We also supplemented our analysis of the reports and statements of these organizations with interviews of officials, at a variety of entities, with knowledge of data-sharing issues. Furthermore, whenever we attribute statements to the scientific community at large, we mean that a National Academies study and at least two of the scientific associations listed above support those statements.

To address the second and third objectives, we identified and reviewed the data-sharing requirements, policies, and practices that exist under the climate change awards—including grants, cooperative agreements, and funded field work proposals—funded by DOE, NASA, NOAA, and NSF. We also interviewed senior officials at DOE, NASA, NOAA, and NSF who direct the climate change research programs. For additional context on how data sharing is carried out, we interviewed managers from data archives that preserve climate change data, including Lawrence Livermore National Laboratory's Program for Climate Model Diagnosis and Intercomparison, the Long-Term Ecological Research Center Network, NOAA's National Climatic Data Center, and Oak Ridge National Laboratory's Carbon Dioxide Information Analysis Center.

To assist in identifying relevant data-sharing requirements, policies, practices, and issues at the four major federal climate change research agencies, we conducted a Web-based survey of all 64 program managers who oversee the climate change research awards at these agencies. We conducted the survey from April 3 to May 3, 2007. To prepare the questionnaire, we pretested potential questions with at least one program manager at each of the four agencies as well as a Senior Earth Scientist with the U.S. Climate Change Science Program to ensure that (1) the questions were clear, (2) terminology was used correctly, (3) questions did not place an undue burden on the respondents, (4) the information was feasible to obtain, and (5) the questionnaire was comprehensive and unbiased. On the basis of feedback from the six pretests we conducted, we made changes to the content and format of some survey questions. The final questionnaire included a combination of open- and closed-ended questions about the data-sharing requirements, policies, and practices at the program manager's agency and specific program.

To ensure an adequate and appropriate response to our questionnaire, agencies provided the names and contact information for climate change

program managers. We contacted all of the program managers in advance to ensure that we had identified the correct respondents. We also sent letters to the program managers informing them of the approximate date the survey would be available, and we then sent an e-mail when the survey was available via the Internet. After the survey had been available for 1 week, we used e-mail and telephone calls to contact the program managers who had not completed their questionnaires. Using these procedures, we obtained a 100-percent response rate. Because this was not a sample survey, there are no sampling errors. However, the practical difficulties of conducting any survey may introduce errors, commonly referred to as nonsampling errors. For example, difficulties in how a particular question is interpreted, in the sources of information that are available to respondents, or in how the data are entered into a database or were analyzed can introduce unwanted variability into the survey results. We took steps in the development of the questionnaire, the data collection, and the data analysis to minimize these nonsampling errors. For instance, a survey specialist designed the questionnaire in collaboration with GAO staff with subject-matter expertise. Further, the draft questionnaire was pre-tested with a number of agency program managers to ensure that the questions were relevant, clearly stated, and easy to comprehend. When the data were analyzed, a second, independent analyst checked all computer programs. In several cases, we contacted respondents to clarify their responses to the questions, but we did not otherwise independently verify the information they provided.

Appendix II: Data-Sharing Policies Applicable to Federally Funded Climate Data

Agency	Climate Change Program	Climate Change Project	Agencywide data-sharing policy ^a	Program or project-specific data-sharing policy ^b
DOE	Atmospheric Radiation Measurement Program			•
	Atmospheric Science Program			•
	Climate Change Prediction Program	Community Climate System Model (CCSM) and Program for Climate Model Diagnosis and Intercomparison (PCMDI)		•
	Terrestrial Carbon Processes	AmeriFlux Network		•
	Program for Ecosystem Research			•
	Integrated Assessment Research Program			•
NASA			•	
	Earth Science Research Program		•	•
	Earth Science Research Program	Research, Education and Applications Solution Network (REASoN)	•	•
	Cryosphere Research Program		•	
	Modeling Analysis and Prediction		•	•
	Ocean Biology and Biogeochemistry Research		•	•
	Atmospheric Composition Modeling and Data Analysis		•	•
	Tropospheric Chemistry	Megacity Initiative: Local and Global Research Observations (MILAGRO)	•	•
	Terrestrial Ecology Program		•	•
	Terrestrial Hydrology Program		•	
	Earth Sciences Program	Land-Cover and Land-Use Change Program	•	
	Upper Atmosphere Research Program		•	
	Radiation Sciences Program		•	•
	Biological Diversity		•	•
NSF			•	
	Ecosystem Studies		•	

**Appendix II: Data-Sharing Policies Applicable
to Federally Funded Climate Data**

Agency	Climate Change Program	Climate Change Project	Agencywide data-sharing policy^a	Program or project-specific data-sharing policy^b
	Sedimentary Geology and Paleobiology		•	•
	Atmospheric Chemistry Program		•	
	National Center for Atmospheric Research	Earth Observing Laboratory	•	•
	Paleoclimate Program		•	
	Climate and Large Scale Dynamics	Climate Variability and Predictability Program	•	•
	Upper Atmospheric Facilities		•	
	Instrumentation and Facilities Program		•	•
	Hydrologic Science		•	
	Education and Human Resources		•	•
	Oceanographic Instrumentation and Technical Services		•	•
	Chemical Oceanography		•	•
	Physical Oceanography		•	•
	Marine Geology and Geophysics		•	•
	Biological Oceanography		•	•
	Global Ocean Ecosystem Dynamics (GLOBEC)		•	•
	Antarctic Glaciology Program		•	•
	Antarctic Organisms and Ecosystems		•	•
	Antarctic Earth Sciences		•	•
	Arctic System Science Program		•	•
NOAA				
	Arctic Research Program			
	Atmospheric Composition and Climate			
	Transition of Research Applications to Climate Services			
	Climate Change Data and Detection			

**Appendix II: Data-Sharing Policies Applicable
to Federally Funded Climate Data**

Agency	Climate Change Program	Climate Change Project	Agencywide data-sharing policy ^a	Program or project-specific data-sharing policy ^b
	Climate Dynamics and Experimental Prediction			
	Climate Observation Program			●
	Climate Variability and Predictability			
	Climate Prediction Program for the Americas			
	Global Carbon Cycle Program			
	Regional Integrated Sciences and Assessments (RISA)			
	Sectoral Applications Research Program			
	Research Cooperative Institute Program			

Source: GAO analysis of survey responses.

Note: The CCSP data-sharing policy, Data Management for Global Change Research Policy Statements, applies to each federal agency.

^aCircle denotes agency-level policies encouraging federally funded researchers to make data available. These policies apply to all of the programs within the agency.

^bCircle denotes program- or project-level policies encouraging federally funded researchers to make data available.

Appendix III: Examples of Data-Sharing Expectations for Different Federally Funded Research Projects

		Expectations for:			
Applicable to projects involving: ^a	Policy	Openness	Accessibility	Timing	Cost
Individual researcher	Grant Policy Manual (NSF) ^b	Researchers are expected to share data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Data should be released in a form that protects privacy and intellectual property; exceptions to the openness expectation may be specified by NSF or requested by grantees.	Grantees are expected to make their data and products widely available and usable.	Researchers are expected to share their data within a reasonable time.	Researchers are expected to make their data available at no more than an incremental cost.
Multiple researchers	AmeriFlux Data Submission Guidelines (DOE) ^c	Researchers should make available the core suite of data.	Researchers should make data available through the central AmeriFlux data repository located at the Carbon Dioxide Information Analysis Center.	Researchers should make data available within 1 year of data collection. Ancillary biological data should be submitted within a reasonable time.	Data are provided free through the central AmeriFlux data repository.
Collection of physical samples	Division of Ocean Sciences Data and Sample Policy (NSF) ^d	Researchers should make all environmental data collected available. Researchers should address alternative strategies for complying with the openness expectation when limitations exist.	Researchers should submit data to designated National Data Centers and samples should be archived at NSF-supported repositories and stored in a manner that preserves the quality of the samples and respects community standards. Where no archive exists for the data, researchers should address alternative strategies to make data available.	Researchers should make all data available as soon as possible, but no later than 2 years after collection; metadata of all marine data collected should be made available within 60 days of the observational period/cruise; for continuing observations, data inventories should be submitted periodically if there is a significant change in such observations.	Data are provided at no more than incremental costs through data archives.

**Appendix III: Examples of Data-Sharing
Expectations for Different Federally Funded
Research Projects**

Applicable to projects involving: ^a	Expectations for:				
	Policy	Openness	Accessibility	Timing	Cost
Collection of satellite data	Data and Information Policy (NASA) ^e	All Earth science data obtained from NASA Earth observing satellites, suborbital platforms and field campaigns, as well as source code for algorithm software, coefficients, and ancillary data, should be made available.	Data will be placed in archives that include accessible information about the data holdings, including quality assessments, supporting relevant information, and guidance for locating and obtaining data.	Data will be made available following the post-launch checkout period. There is no period of exclusive access to NASA Earth science data.	NASA will charge for distribution of data no more than the cost of dissemination. In cases where such dissemination cost would unduly inhibit use, the distribution charge will generally be below that cost.
Modeling data	Community Climate System Model (DOE/NSF) ^f	The CCSM project is committed to making available the results from the model runs and the scientific data generated in research activities. Short models runs made to examine specific model behavior, validate a port of the code to a new platform, or verify model functionality do not need to be made available.	CCSM model data should be retained in public archives for a period of 10 years from the date of the end of the simulation.	Researchers are entitled to a proprietary period during which they can publish results but are encouraged to share their data prior to the release deadlines; data shall become public when a paper has been submitted or 1 year after the end of the simulation, whichever comes sooner.	Data will be made available for users who are not CCSM collaborators at the marginal cost of making and shipping the copies; however, for large data orders, special arrangements may be made.

Source: GAO analysis.

^aThese policies are generally not exclusively relevant to the particular types of research projects noted here. For example, a policy labeled as applying to "individual researcher" in this column could also be applicable to projects with multiple researchers.

^bSee section 734 of NSF's Grant Policy Manual, available at http://www.nsf.gov/pubs/manuals/gpm05_131/gpm05_131.pdf.

^cSee the Oak Ridge National Laboratory's AmeriFlux Network Data Guidelines, available at <http://public.ornl.gov/ameriflux/data-guidelines.shtml>. AmeriFlux is a project that helps coordinates regional and global analysis of observations from micrometeorological tower sites.

^dSee the General Data Policy and Sample Policy sections of the NSF's Division of Ocean Sciences Data and Sample Policy, available at <http://www.nsf.gov/pubs/2004/nsf04004/start.htm>.

^eSee NASA's Data and Information Policy (part of their Earth Science Reference Handbook), available at <http://science.hq.nasa.gov/research/daac/datainfo/policy.pdf>.

^fSee CCSM's Data Management Plan, available at <http://www.cesm.ucar.edu/experiments/data.mgmt.plan.050803.html>.

Appendix IV: Governmentwide Climate Change Data-Sharing Policy

This appendix presents an annotated copy of an interagency policy, the Data Management for Global Change Research Policy Statements. This policy was written by the U.S. Global Change Research Program Data and Information Working Group and was later endorsed by the U.S. Climate Change Science Program. The italicized text provides comments on key issues in the policy, such as how and what data to share. The policy also includes a suggested data product requirement and compliance guidelines, about which we do not comment.

U.S. Global Change Research Program Data and Information Working Group Data Management for Global Change Research Policy Statements

Background

In 1991 the Executive Office of the President, Office of Science and Technology Policy issued the following data management for global change policy statements. The overall purpose of these policy statements was to facilitate full and open access to quality data for global change research.

GAO Comment

The policy statements reflect the scientific community's belief that data-sharing policies should address what, how, and when data are to be shared, as well as the cost of making data available.

Though the policy statements are applicable to all federally funded climate change research, they are not legal requirements, as the statements discuss under the Compliance section below. Senior officials at the four major climate change research agencies we reviewed told us that their data-sharing policies and practices follow the principles of these statements.

They were prepared in consonance with the goal of the U.S. Global Change Research Program and represent the U.S. Government's position on the access to global change research data.

Applicability

1. Federally funded data significantly related to the USGRP that includes:

A. Data resulting from observations, the application of algorithms to data to produce new data, and from the data output of models.

B. Data resulting from agency funding in whole or in part of inhouse activities or of cooperative, grant, and contracted activities. Included is the data an agency purchases of data from outside the government to meet its needs*.

GAO Comment

This section addresses what data to share by defining the information and materials expected to be shared.

(* Such an inclusion of purchased data is included in the 2001 NAS report “Resolving Conflicts Arising from the Privatization of Environmental Data.”)

2. While it is hoped that these guidelines would be as broadly applied as possible, their intent is primarily focused on providing guidance for when new data is being obtained and made available or when existing data because of technology or other changes needs to be reformatted or have other such changes.

Guidelines and Their Application

POLICY STATEMENT 1. The U.S. Global Change Research Program requires an early and continuing commitment to the establishment, maintenance, validation, description, accessibility, and distribution of high-quality, long-term data sets.

GAO Comment

This policy statement addresses how and what data to share. The statements follow the scientific community’s belief that data should be made available via unrestricted archives.

Since 1994 the USGCRP has managed a Web page, the Global Change Data and Information System, GCDIS, that helps users find the largest amount of USGCRP related data of any site in the world. In 1999, it also became the largest site for data policy information. This site is at <http://www.globalchange.gov/>

A. Applicable agency data should be made readily accessible to potential users:

Minimum application – All such data used in openly available publications, reports, and analyses.

GAO Comment

The statements follow the scientific community's belief that, at a minimum, all data necessary to support researchers' major published results should be made available. It may not always be possible or appropriate to share all data, such as in modeling research where only certain data outputs are relevant to the larger scientific community.

Desired application – All such significant data produced.

B. Applicable agency data should be made available via the Web:

Minimum application – All such data used in openly available publications, reports, and analyses that are in digital form.

Desired application – All such significant data that's openly available.

C. Applicable data made available on the Web should be described with each data set having:

Minimum application – A citation similar to those used for citing publications in research journals and in use for data sets by the USGCRP since 1997.

Desired application – A citation plus a data set description that (1) can be readily found and is adequate for users to be able to both understand the applicability of the data to their needs and its proper use and (2) meets at least the minimum requirements for inclusion in the Global Change Master Directory, GCMD, and is so identified to the GCMD.

POLICY STATEMENT 2. Full and open sharing of the full suite of global data sets for all global change researchers is a fundamental objective.

GAO Comment

This policy statement addresses what data to share. The statements follow the scientific community's belief that data sharing should be full and open.

This objective has since 1991 been repeatedly urged and defended from compromise by the National Academy of Science, NAS. The concept has also been widely adopted and applied both nationally and internationally. After reviewing all these implementation actions, the NAS recommended the following single definition "Full and open availability is defined as being available without restriction, on a non-discriminatory basis, for no

more than the cost of reproduction and distribution.” It combines elements of this Statement with those of Statement 6 was adopted by the USGCRP in 1997.

A. Full and open access to agency data sets should be provided to:

Minimum application – All agency data related to the USGCRP that’s made generally available.

GAO Comment

As noted above, some data are not appropriate for sharing. Full and open, in these policy statements, does not necessarily mean that every data item or iteration of data analysis must be made available. What data to share are generally more specifically defined by relevant agency, program, or project policies.

Desired application – All agency data that’s made generally available.

POLICY STATEMENT 3. Preservation of all data needed for long-term global change research is required. For each and every global change data parameter, there should be at least one explicitly designated archive. Procedures and criteria for setting priorities for data acquisition, retention, and purging should be developed by participating agencies, both nationally and internationally. A clearinghouse process should be established to prevent the purging and loss of important data sets.

GAO Comment

This policy statement addresses how to share data. The statements follow the scientific community’s general belief that, when an appropriate archive exists, data should be shared via unrestricted archives. The lack of appropriate archival infrastructure can be an obstacle to data sharing.

The Federal requirement for providing adequate notice when agencies purge significant data and information products is called for in OMB Circular A-130 of 1997.

A. The USGCRP should be notified of any agency plans to purge data significantly related to the USGCRP program so an interagency process can determine the necessary remedial actions, if any.

Minimum application – Notification at least six months prior to the data being purged, or as soon as the agency’s intent seems likely, whichever is shorter.

Desired application – Notification as soon as the data purging is being seriously considered by an agency.

(It should be noted that these guidelines apply equally well in normal times and in abnormal times, such as after the 9/11/01 event.)

POLICY STATEMENT 4. Data archives must include easily accessible information about the data holdings, including quality assessments, supporting ancillary information, and guidance and aids for locating and obtaining the data.

GAO Comment

This policy statement addresses how and what data to share. The statements reflect the belief that data should be shared with their metadata, i.e., information needed to understand the content, quality, and condition of the raw data.

A. For the applicable data that agencies make available, an assessment of its quality is needed to help assure its proper use.

Minimum application – Identification of the source of the data so the user has a place to check on its quality.

Desired application – Identification of the data's quality sufficient to assure its proper use and make unlikely its improper use.

(The requirement for the identifying the quality of data made available is contained in OMB's "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Provided by Federal Agencies" issued in 2001.)

B. For the applicable data that agencies make available there should be the ability to be responsive to users questions relative to its use.

Minimum application – A means for the user to identify the source of the data, i.e. the specific person or organization responsible.

Desired application – Identification of a person or organization that will be responsive to a user's requests for help.

C. To maximize the ability of users to use the applicable data made available, the vision is to have data from different sources be able to be seamlessly used with data taken by other means, from different sources, and measuring other parameters. That is, have full interoperability.

Minimum application – Enough data is provided with a data set so its user can make it interoperable with other data sets.

Desired application – Meets the preceding “Minimum application” and the data set has at least spatial and temporal interoperability with the other such interoperable data within the USGCRP.

POLICY STATEMENT 5. National and international standards should be used to the greatest extent possible for media and for processing and communication of global data sets.

GAO Comment

This policy statement addresses how to share data.

A. In 1994 Executive order 12906 created the National Spatial Infrastructure, NSDI, and OMB Circular A-16 its Federal Geographic Data Committee, FGDC, management structure. For all geospatial data, agencies are must have compatibility with their data documentation standards. The FGDC actively tries to assure their standards are compatible with international standards.

Minimum Application – All applicable data when new data is being obtained and made available or when existing data because of technology or other changes needs to be reformatted or have other such changes.

Desired Application – All applicable data.

B. In 1995 the parent group of the USGCRP, OSTP’s Committee on Environment and Natural Resources, instructed its participating agencies to have their individual data and information access and search systems be in compliance with the American National Standards Institute, ANSI, Z39.50 10162/10163 open standards for information search and retrieval.

Minimum Application – All applicable data when new data is being obtained and made available or when existing data because of technology or other changes needs to be reformatted or have other such changes.

Desired Application – All applicable data.

POLICY STATEMENT 6. Data should be provided at the lowest possible cost to global change researchers in the interest of full and open access to data. This cost should, as a first principle, be no more than the marginal

cost of filling a specific user request. Agencies should act to streamline administrative arrangements for exchanging data among researchers.

GAO Comment

This policy statement addresses the cost of sharing data. The statements follow the scientific community's belief that data should be made available at no more than the marginal cost of reproduction and distribution.

The Federal requirement for charging users no more than the marginal cost of servicing their request is called for in OMB Circular A-130 of 1997.

Minimum Application – All applicable data.

POLICY STATEMENT 7. For those programs in which selected principal investigators have initial periods of exclusive data use, data should be made openly available as soon as they become widely useful. In each case the funding agency should explicitly define the duration of any exclusive use period.

GAO comment

This policy statement addresses when data should be shared. The statements follow the scientific community's belief that data should generally be made available immediately or after a limited proprietary period that allows researchers to complete their initial analysis and publish their results. The scientific community generally recognizes the need for researchers to have a limited period of exclusive access to their data to allow for analysis, interpretation, and peer review that normally precedes public disclosure. However, the length of such a period may be determined by the type of research.

To meet this need, in 1997 the USGCRP endorsed the following grant language for use by its participating agencies.

Suggested Data Product Requirement for Grants, Cooperative Agreements, and Contracts

Describe the plan to make available the data products produced, whether from observations or analyses, which contribute significantly to the <grant's> results. The data products will be made available to the <grant official/contracting officer> without restriction and be accompanied by comprehensive metadata documentation adequate for specialists and non-specialists alike to be able to not only understand both how and where the data products were obtained but adequate for them to be used with confidence for generations. The data products and their metadata will be provided in a <standard> exchange format no later than the <grant's> final report or the publication of the data product's associated results, whichever comes first.

Minimum Application – All such applicable data identified as important to the USGCRP.

Desired Application – All such applicable data.

Compliance

While these guidelines themselves are not requirements on the agencies, many result from Federal requirements that do require agency compliance. Rather the guidelines' goal is to help provide guidance to the agencies on how best to meet the needs of users for USGCRP related data within their resources. As such, to help users of a particular data set made available by an agency readily understand the degree to which it meets the guidelines, as well as to recognize the efforts an agency to meet these guidelines:

1. Provided a data set meets all of the Federal requirements and at least all the minimum levels of guideline application the agency should add a single asterisk at the end of the data set's citation.
2. Provided a data set meets all of the Federal requirements and all the desired levels of guideline application - the agency should add two asterisks at the end of the data set's citation.

For broader compliance than for selected individual data sets:

Minimum compliance - Endorsement of these guidelines at the highest appropriate level in the agency.

Desired compliance - Incorporation of these guidelines into the data management policies of the highest appropriate level in the agency.

Appendix V: Comments from the Department of Energy



Department of Energy
Washington, DC 20585

SEP 07 2007

John Stephenson
Director, Natural Resources and Environment
Government Accountability Office
441 G St., NW, Room 2T23A
Washington, DC 20548

Dear Mr. Stephenson:

Thank you for the opportunity to comment on the draft Government Accountability Report, entitled "Climate Change Research: Agencies Have Data-Sharing Policies but Could Do More to Enhance the Availability of Data from Federally Funded Research." We appreciate your efforts to assist in assuring that the data and information on Climate Change is available to the public. The Department and the partner agencies in the Climate Change Science Program consider this a priority objective. Overall, we agree with your principle finding and recommendation that the agencies consider additional processes to enhance data availability.

Below, please find additional general and specific comments for your consideration:

General Comment:

The report should clearly define what is meant by "data." There are several brief references to both electronic-type data and actual field samples. For example, on page 28 the discussion implies that both are included when referring to data throughout the report. In general, researchers do not think of samples collected in (or from) experiments as data but as project resources whereas the data are what results from the analysis of the samples. A clear definition of "data" is needed up front.

Sharing of electronic data, while complicated and potentially expensive in its own right, is far less complicated than the sharing of actual research samples which are also finite in the amount of materials actually available. If the report intends to include samples as well, then a discussion is needed of how much of a finite sample should be available and how much should a researcher be able to keep for their own future use. The amount will depend on what future analysis might be done on the samples which will be virtually impossible to predict. Such research samples should not be considered the same as data.

Specific Comments:

Page 3 reference to ice cores is unclear – Does GAO mean to say Antarctic instead of Arctic?



Printed with soy ink on recycled paper

Page 6, 1st full paragraph, "...the scientific community has traditionally only rewarded publication." COMMENT/CLARIFICATION: It is not just the scientific community that rewards publication, but also institutions, agencies and performance reviews that place a higher value on publication of original results, and give lesser credit to data management functions. The reward matter was discussed in broader terms in the body of the report, but this condensed version in the "In Brief" section seems to be an over simplification.

Highlights page, page 6, page 25, and page 26) the report says something such as "agencies do not monitor whether researchers make data available." This is an over simplification. Several programs, such as the Department of Energy ARM and AmeriFlux, do monitor if data is placed in a public archive, and also monitor if the data is accessed by individuals outside the program. DOE recommends that such statements be revised along the lines of "agencies do not ROUTINELY monitor whether researchers make data available FROM ALL THEIR RESEARCH PROGRAMS."

On page 8, the report says that "...agencies request written reviews...to assess the scientific merit of proposals in SOME cases". For the DOE climate research, ALL proposals have written reviews for scientific merit. The last phrase might therefore be changed from "in some cases" to "in most or all (depending on agency) cases."

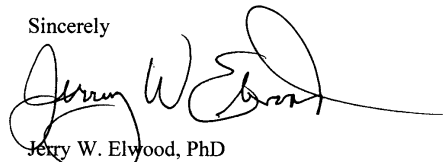
Page 24, paragraph in middle of the page: The last sentence is incomplete; it should read, "...Carbon Dioxide Information Analysis Center, which preserves data from researchers collaborating in the AmeriFlux Network and other Climate Change Programs." COMMENT: Only a fraction of the \$2.7 million supports the AmeriFlux data base, and the total budget supports other data management activities. Note there is no "and" in "Oak Ridge National Laboratory's Carbon Dioxide Information Analysis Center" (on page 35, as well).

In several places the report states that data archiving and sharing is not part of the reward structure for scientists. While it is perhaps not "generally" a "significant" part of the reward system, there are cases where important peer recognition goes along with the dissemination of long-term data records. The statement "...agencies and the scientific community expect researchers to both publish their results and make underlying data available but researchers have traditionally ONLY been rewarded for publication" (page 29) could more accurately be stated along the lines of "...agencies and the scientific community expect researchers to both publish their results and make underlying data available but researchers have traditionally been rewarded MAINLY for publications." Similarly, the Highlights page (near the bottom) could say "...but preparation of data for others' use is not A SIGNIFICANT part of this reward structure."

We look forward to working with our partner agencies as we consider additional opportunities and processes to make climate change research data available to the research community.

If you have any questions, please contact me at 301-903-3281.

Sincerely

A handwritten signature in black ink, appearing to read "Jerry W. Elwood", with a long horizontal flourish extending to the right.

Jerry W. Elwood, PhD
Acting Associate Director of Science
for Biological and Environmental Research

Appendix VI: Comments from the National Aeronautics and Space Administration

National Aeronautics and
Space Administration
Office of the Administrator
Washington, DC 20546-0001



September 17, 2007

Mr. John Stephenson
Director, Natural Resources and Environment
United States Government Accountability Office
Washington, DC 20548

Dear Mr. Stephenson:

NASA appreciates the opportunity to comment on your draft report entitled, "Climate Change Research: Agencies Have Data-Sharing Policies but Could Do More to Enhance the Availability of Data from Federally Funded Research" (GAO-07-1172).

In the draft report, GAO makes four recommendations regarding the availability of data from Federally funded climate change research. However, of the four recommendations made, only three (recommendations 2 – 4) are addressed to the NASA Administrator:

Recommendation 2: To ensure that the agencies maximize opportunities to make data available in a manner useful to other researchers, we recommend that the Secretary of Energy, the Administrator of NASA, the Secretary of Commerce and the NOAA Administrator, and the Director of the National Science Foundation consider the following actions:

- Develop mechanisms for agencies to be systematically notified when data have been submitted to archives, so that agency officials have current information about the extent of data availability in order to adjust data-sharing policies over time to best meet the needs of researchers and the communities that use their data.

Response: NASA concurs with this recommendation and already has such mechanisms in place for its archives. NASA officials (and the public at large) have current information about the extent of data availability and the ability to adjust data-sharing policies over time to best meet the needs of researchers and the user communities that use their data.

Recommendation 3: Use the grants process, where their program offices are not currently doing so, to facilitate further data sharing by: (1) evaluating researchers' data-sharing plans as part of the grant review process, and (2) using evidence of researchers' past data sharing practices to make future award decisions. The use of such

criteria in the grant review process should be clearly conveyed to researchers before they submit research proposals and after award decisions have been made.

Response: NASA concurs with the intent of this recommendation and believes NASA policies facilitate data sharing. NASA employs “full and open exchange” data policy for its satellite data and standard products holdings.

For grantees, NASA’s guidelines with regard to releasing data and results derived through its research awards can be found in the ROSES Guidebook for Proposers. These guidelines state that NASA may, where appropriate, require that any data obtained through an award be deposited in an appropriate public data archive as soon as possible after calibration and validation.

Recommendation 4: To ensure that researchers make climate change data available to other researchers, we recommend that the Secretary of Energy, the Administrator of NASA, the Secretary of Commerce and the NOAA Administrator, and the Director of the National Science Foundation:

- Evaluate whether additional strategies are warranted to facilitate the permanent archiving of relevant data, which may include leveraging existing resources, devoting a greater portion of data collection funds to archiving activities, or working with existing entities such as the National Science and Technology Council’s Interagency Working Group on Digital Data, to develop additional data archives.

Response: NASA concurs with this recommendation. NASA’s Earth Science Program systematically evaluates the demand for data products by the science community and users. NASA analyzes individual data collections and develops the best methodologies for archival and distribution for these collections. NASA then focuses on developing the most cost-effective system for data archival and distribution of service to the science community and the Nation. NASA’s current capacity for data archive and distribution is sufficient for all relevant data for the foreseeable future.

Again, thank you for the opportunity to review and comment on this draft report and for the critical insight it provides. If you have any questions, please contact Michael Luther on (202) 358-7226.

Sincerely,

A handwritten signature in black ink, appearing to read 'Shana Dale', with a stylized flourish at the end.

Shana Dale
Deputy Administrator

Appendix VII: Comments from the Department of Commerce, for the National Oceanic and Atmospheric Administration



THE SECRETARY OF COMMERCE
Washington, D.C. 20230

September 10, 2007

Mr. John Stephenson
Director, Natural Resources
and Environment
U.S. Government Accountability Office
441 G Street, NW
Washington, D.C. 20548

Dear Mr. Stephenson:

Thank you for the opportunity to review and comment on the Government Accountability Office's draft report entitled *Climate Change Research: Agencies Have Data-Sharing Policies but Could Do More to Enhance the Availability of Data from Federally Funded Research* (GAO-07-1172). On behalf of the Department of Commerce, I enclose the National Oceanic and Atmospheric Administration's programmatic comments to the draft report.

Sincerely,

A handwritten signature in black ink, appearing to read "Carlos M. Gutierrez", is written over a printed name.

Carlos M. Gutierrez

Enclosure

Department of Commerce
National Oceanic and Atmospheric Administration
Comments on the Draft GAO Report Entitled
“Climate Change Research: Agencies Have Data-Sharing Policies but Could Do More to
Enhance the Availability of Data from Federally Funded Research”
(GAO-07-1172/September 2007)

General Comments

The Department of Commerce appreciates the opportunity to review this report. The report does a fair job in assessing the National Oceanic and Atmospheric Administration’s (NOAA’s) data-sharing policies for climate change research. The discussion of practices to encourage data-sharing, and the associated obstacles and disincentives, is accurate and balanced.

NOAA Response to GAO Recommendations

The draft GAO report states, “To assist federal agencies sponsoring climate change research in better ensuring the availability of data from federally funded research, we are making the following four recommendations.”

Recommendation 1: *“To ensure that researchers receiving federal funds to conduct climate change research understand NOAA’s expectations for data sharing, we are recommending that the Secretary of Commerce and the NOAA Administrator: Develop a set of written guidelines or use existing government-wide guidelines, such as those endorsed by the Climate Change Science Program, to clearly inform researchers of NOAA’s general expectations for data sharing.”*

NOAA Response: NOAA agrees with this recommendation. There are extensive existing guidelines available for use by NOAA to mirror or enhance. Therefore, through the Climate Change Science Program’s Science Advisory Board, NOAA plans to address this recommendation and take the necessary steps (i.e., develop new guidelines or enhance existing guidelines) to more clearly inform researchers of NOAA’s data sharing expectations.

Recommendation 2: *“To ensure that the agencies maximize opportunities to make data available in a manner useful to other researchers, we recommend that the Secretary of Energy, the Administrator of NASA, the Secretary of Commerce and the NOAA Administrator, and the Director of the National Science Foundation consider the following actions: Develop mechanisms for agencies to be systematically notified when data have been submitted to archives, so that agency officials have current information about the extent of data availability in order to adjust data-sharing policies over time to best meet the needs of researchers and the communities that use their data.”*

NOAA Response: NOAA agrees with this recommendation and believes agency officials should have current information about the extent of data availability. Given other agencies are also involved, NOAA will do its part and consider developing a mechanism to systematically notify others when data have been archived. As part of this process, however, several things will need to be considered, including: (1) funding implications; (2) the need for standard reporting categories to facilitate data collection (rather than the use of the numerous agency-unique reporting forms and systems); and (3) the benefits of a mechanism which would enhance the ability of both the public and agencies to organize the data.

Recommendation 3: *"To ensure that the agencies maximize opportunities to make data available in a manner useful to other researchers, we recommend that the Secretary of Energy, the Administrator of NASA, the Secretary of Commerce and the NOAA Administrator, and the Director of the National Science Foundation consider the following actions: Use the grants process, where their program offices are not currently doing so, to facilitate further data sharing by: (1) evaluating researchers' data-sharing plans as part of the grant review process, and (2) using evidence of researchers' past data-sharing practices to make future award decisions. The use of such criteria in the grant review process should be clearly conveyed to researchers before they submit research proposals and after award decisions have been made."*

NOAA Response: NOAA agrees with this recommendation and believes taking the recommended actions will emphasize to researchers applying for grants NOAA's commitment to data-sharing. NOAA plans to include past, present, and future data sharing as an evaluation factor in its grant opportunity announcements. NOAA will also consider including additional rating factors in assessing grant proposals, as well as the inclusion of additional requirements in grant awards for data sharing by the awardee to encourage good data sharing practices. Moving forward, NOAA will pursue avenues to more clearly convey the importance of these factors to researchers prior to the submission of research proposals and after award decisions have been made.

Recommendation 4: *"To ensure that researchers make climate change data available to other researchers, we recommend that the Secretary of Energy, the Administrator of NASA, the Secretary of Commerce and the NOAA Administrator, and the Director of the National Science Foundation: Evaluate whether additional strategies are warranted to facilitate the permanent archiving of relevant data, which may include leveraging existing resources, devoting a greater portion of data collection funds to archiving activities, or working with existing entities such as the National Science and Technology Council's Interagency Working Group on Digital Data, to develop additional data archives."*

NOAA Response: NOAA agrees with this recommendation. Permanent archiving is very important and should be incorporated into the strategic plans of archiving facilities. NOAA will consider whether additional strategies are warranted to facilitate the permanent archiving of data, as well as the additional resource requirements to support such strategies.

Recommended Changes for Factual/Technical Information

Page 12, First Paragraph:

The need for complete metadata should be identified.

We suggest adding at the end of this paragraph, "For example, in all cases sufficient metadata, such as data set descriptions, should be provided so the data can be found and its suitability for use determined. This need recognizes most data will be archived, and other researchers will need to search the archives to identify useful data."

Appendix VIII: GAO Contact and Staff Acknowledgments

GAO Contact

John Stephenson, 202-512-3841 or stephensonj@gao.gov

Staff Acknowledgments

In addition to the contact person named above, Diane Raynes (Assistant Director); Kyle Browning; Kate Cardamone; John Delicath; Carolyn Garvey; Richard Johnson; Lynn Musser; and Katherine M. Raheb made key contributions to this report.

GAO's Mission

The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's Web site (www.gao.gov). Each weekday, GAO posts newly released reports, testimony, and correspondence on its Web site. To have GAO e-mail you a list of newly posted products every afternoon, go to www.gao.gov and select "E-mail Updates."

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are \$2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. Government Accountability Office
441 G Street NW, Room LM
Washington, DC 20548

To order by Phone: Voice: (202) 512-6000
TDD: (202) 512-2537
Fax: (202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm

E-mail: fraudnet@gao.gov

Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

Gloria Jarmon, Managing Director, JarmonG@gao.gov, (202) 512-4400
U.S. Government Accountability Office, 441 G Street NW, Room 7125
Washington, DC 20548

Public Affairs

Susan Becker, Acting Manager, BeckerS@gao.gov, (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, DC 20548