
DATA MANAGEMENT FOR GLOBAL CHANGE RESEARCH POLICY STATEMENTS

July 1991

The overall purpose of these policy statements is to facilitate full and open access to quality data for global change research. 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.

- The 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.
 - Full and open sharing of the full suite of global data sets for all global change researchers is a fundamental objective.
 - 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.
 - 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.
 - National and international standards should be used to the greatest extent possible for media and for processing and communication of global data sets.
 - 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.
 - 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.
-

ANNEX

DATA MANAGEMENT FOR GLOBAL CHANGE RESEARCH POLICY STATEMENTS

July 1991

The 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.

Agencies involved in global change research noted that inadequate attention has often been given in the past to the creation and maintenance of quality long-term data sets. Often this neglect was attributed to relatively lower priority given to long-term data management compared with initial data collection and analysis, with a concomitant lack of resources for the longer term effort. The Interagency Working Group on Data Management for Global Change (IWGDMGC), which assisted in development of these policy statements, pointed out that the long-term cost of maintaining large volumes of data can be significant and suggested that the required resources for this purpose must be committed at the start of data collection projects.

Furthermore, the proper preparation, validation, description, and care of data sets is critical to their use by the widest possible scientific community. Those not involved in the initial data collection and processing must be able to easily determine how data have been collected, calibrated, validated, and otherwise transformed. This may include the development of community-consensus algorithms and instructional efforts to ensure that potential users are aware of data availability.

In some cases the responsibility for establishing and maintaining global change research data sets may be shared by agencies other than the originators of the data collection efforts. Plans must be developed as part of the overall project to ensure that the investment in data collection is enhanced and expanded by adequate long-term data management practices.

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

Federal agencies have different data distribution practices affecting global change research data. The IWGDMGC proposes establishing a fundamental objective of full and open sharing of the full suite of global data sets for all

global change researchers. Data sets should be made available in a timely manner, but the definition of timeliness is left as a responsibility of the funding agencies involved. As data are made available, global change researchers should have full and open access to them without restriction on research use.

Global change researchers include those in academic, industry, government, and non-government sectors conducting both basic and applied research.

The global change research data sets contain data of potential usefulness to a competitive U.S. economy for industrial applications and improved environmental management. As required by appropriate public law, global change research agencies will develop plans for commercial access to the global change databases.

To accomplish this objective, data must be submitted to archives, and information about data sets must be created and made available as well. The access policies for these archives should encourage the widest possible use of global change research data in meeting the objectives of the Global Change Research Program.

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.

The agency representatives noted that data sets representing some of the measurement parameters important to global change research do not presently have an archive "home". Many of the biological parameter were cited as an example.

This policy statement is meant to emphasize the responsibility of data collecting and producing agencies to identify suitably supported, long-term archives for all data sets important to global change research, make arrangements for those archives to acquire the data sets and related information, and make them available for open research use. This principle is not meant to exclude distributed or multiple archives where appropriate for particular data sets, but to establish, at a minimum, one explicitly designated archive for each global change research parameter.

In light of the high cost of long-term data maintenance, the IWGDMGC recommends the establishment of specific criteria and procedures for setting priorities for data acquisition, retention, and purging. Some data may not be worth retaining on a long-term basis due to poor quality or other considerations such as cloud cover. However, a mechanism should be developed to ensure that the research community is consulted prior to decisions that result in data loss. This includes the opportunity for a new organization to assume responsibility for maintaining data sets no longer given a high priority by the original archival agency.

This consultative and clearinghouse process should include international as well as national organizations. This might provide a reciprocal opportunity for US agencies to participate in decision making by non-US agencies that hold data of interest to the US.

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.

Archive data should include supporting information sufficient to permit its effective use by researchers not familiar with the original data collection project or the particular instrument making the measurements. One limitation on using existing data sets by scientists involved in global change research is the difficulty encountered in identifying what data exist, how to access them, and what the real meaning is of the information contained in such data sets. In the absence of supporting documentation on instrument calibration, validation campaigns, and other ancillary information, full evaluation and application of existing data can be limited. The repositories for global change research data sets must recognize their obligation to obtain or develop full accompanying information for all global change research data holdings and make the data and the supporting information easily available. This requires a well-conceived directory, catalog, and inquiry system.

Peer review is one important mechanism for establishing and documenting data quality. The IWGDMGC, however, notes that peer review may not always be necessary before data release. What is essential is that data be well enough documented to ensure that users can understand what they are getting.

Work under way through the IWGDMGC to establish an interagency Global Change Master Directory (GCMD), and eventually a more comprehensive Global Change Data and Information System (GCDIS), should contribute to accomplishing this objective. Through linking individual agency directories,

users will be able to obtain information about existing data holdings anywhere in the interagency complex without having to separately contact each individual agency. Once data of interest are located, the user can then proceed to obtain the data of interest from the archive where the data reside.

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

Use of standard media, and processing and communications protocols and procedures, aims at making data accessible in a "vendor-independent" environment. The diverse user community has investigated in many different types of data analysis systems. To the extent possible, through standards and protocols, users should be able to obtain, read, and process data without needing to design or purchase data-specific hardware, software, and systems.

Much progress has been made through national and international standards organizations, some of which address very broad areas of application and others which are more discipline or application specific. For example, the International Standards Organization has an Open Systems Interface protocol with seven different layers of interconnection for communications systems. This work is stimulated by many industries and potential users far beyond the global change research community. The Committee on Earth Observations Satellites is an international organization comprising satellite operators and is developing standard formats for user products from specific types of sensors on remote sensing satellites. These efforts and others should be encouraged and supported by IWGDMGC agencies, and the resulting standards and protocols should be used in global change research projects.

The critical objective of standards use is to ensure the widespread availability and use of data. The emphasis is on ensuring that data sets are available to users in standard formats and through agreed communications protocols where applicable, not necessarily that the internal details of individual agency data handling and data archiving systems be common.

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.

Agencies are governed by a wide variety of policies and practices in data charging and pricing. For researchers (defined differently at different agencies) data are usually, but not always, provided either free of cost or at the marginal cost of reproduction and distribution.

There was recognition by the IWGDMGC that charging the marginal cost of reproduction and distribution can be an effective tool for managing requests for large data sets without restricting access. It also permits data distribution agencies to support widespread data use without adverse budget impacts. For small data sets and those accessed infrequently, the administrative burden of marginal cost recovery may outweigh the benefits of charging such costs, and data may be more efficiently provided at no cost. The essential principle is that research users should not be subject to commercial, profit-based pricing for data sets to be used in support of publicly-sponsored global change research.

In addition to the charging practices, administrative arrangements should be streamlined to facilitate data access and exchange. The Global Change Data and Information System development effort is beginning to address these issues.

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.

The agreed objective of this data policy statement is to facilitate full and open access to quality data on a timely basis. Although some data are made available as soon as the data are collected, some agencies provide initial periods of exclusive data use for selected investigators so that data evaluation and validation can be accomplished before general release. Data are not always fully documented and useful during the initial data collection and analysis period, and the need for flexibility in data release was recognized by the IWGDMGC.

Deciding when data became widely useful is the responsibility of the funding agency, which should explicitly define the period of restricted access, if any. In the past, some Principal Investigators have retained data for indefinite periods, and this has inhibited their widespread use. This practice should be eliminated through active consideration of the tradeoffs between widespread distribution of data sets and the need to assure data quality and validity. The guiding principle is that as soon as data might be useful to other researchers the data should be released, along with documentation which can be used by the other

researchers to judge data quality and potential usefulness. In this way users can determine for themselves if they want to proceed with data of questionable quality or wait for additional developments.

The U.S. Global Change Research Program (USGCRP) was conceived and developed to be policy-relevant, and hence, to support the needs of the United States and other nations to address significant uncertainties in knowledge concerning the natural and human-induced changes now occurring in the Earth's life-sustaining environmental envelope.

Prepared for the U.S. Global Change Research Program