United States Department of the Interior

U. S. GEOLOGICAL SURVEY
Reston, VA 20192

In Reply Refer To:
Mail Stops 411, 412, 415

December 5, 2014

OFFICE OF GROUNDWATER TECHNICAL MEMORANDUM 2015.02
OFFICE OF SURFACE WATER TECHNICAL MEMORANDUM 2015.01
OFFICE OF WATER QUALITY TECHNICAL MEMORANDUM 2015.01

SUBJECT: Policy and Guidelines for Archival of Surface-Water, Groundwater, and Water–Quality Model Applications

BACKGROUND

As a public agency, the U.S. Geological Survey (USGS) is responsible for making the results of its scientific investigations widely available to the public in the form of timely, technically-sound, and professionally-presented information products. This memorandum addresses the archival of model applications (model archives) completed as part of USGS scientific research and investigations. The Water Mission Area (WMA) model archival procedures described herein are intended to help the USGS more effectively fulfill policy and statutory requirements.

The National Archives and Records Administration (NARA) requires archival of data and information under the Federal Records Act. The WMA conforms to this statutory responsibility with Scientific Records Disposition Schedules for proper archival of project information. Specifically, "Research and Investigative Project Case file 1400-02" establishes the requirement for archiving all records associated with computer models. Model archives also preserve information consistent with guidance in OSTP Memorandum dated February 22, 2013 on Increasing Access to the Results of Federally Funded Scientific Research and OMB Memorandum M-13-13 on Open Data Policy-Managing Information as an Asset. Model archives help to efficiently meet requests made under the Freedom of Information Act and Information Quality Act. Model archives also preserve highly valuable information for future use and application to water monitoring, assessment, and research activities.

The archiving (and distribution) of model software developed and intended for public distribution and application is covered under separate policy (Water Resources Discipline Policy Memorandum 2009.01).

Since 1975, the Office of Groundwater (OGW) has provided leadership in the development of policies regarding documentation and archival of numerical groundwater model applications.

The evolution of policy to current standards can be found in eight technical memoranda from OGW (see REFERENCES section).

PURPOSE

The purpose of this memorandum is to create common requirements and policies across the WMA and WMA-funded activities to maintain digital model archives. This policy applies to all models that are a significant part of investigations reported in USGS interpretive publications, or other information products.

A model is defined herein as any mathematical representation of a system that is used to study the effects of individual system components and their interactions, or predict future system behavior. Examples of the types of models included in this policy are hydraulic, hydrologic, statistical, geostatistical, physical (such as sediment and heat), chemical, hydrobiological, and biophysical models. The model application must be archived if the results of the model form the basis for conclusions published in USGS reports or other information products.

POLICY

All published model applications must be documented in a manner that facilitates the reproducibility of the results presented in the published information product. The policy and guidelines described herein applies to all WMA Programs, Water Science Centers (WSC), Technical Offices, and National Research Program (NRP) Centers collectively referred to as "Centers" in this memorandum.

An electronic model archive is required for those model applications for which the required level of documentation cannot be accommodated in the information product used to report on the modeling. The electronic model archive will be created and stored in the Center which was primarily responsible for the modeling. The archive is required to: (1) support and validate the results in published reports, (2) assure that working versions of all models are available for future scientific use, and (3) assure that the data are available to the public when requested.

An electronic model archive is not required if one or more of the following conditions apply.

1. The results are from modeling applications that are well-described in the information product and the publication contains enough information to reproduce the model results. The information required to reproduce the modeling application may be contained in supplemental information included in the publication.

2. The results are from routinely used models for established laboratory and field procedures that are documented by other means (for example, simple regression models derived from calibration data that is stored elsewhere and archived as described in NARA Scientific Disposition Schedules).

3. The results are described in an outside publication and none of the USGS coauthors were involved in the model development or application.

The verification of the reproducibility of the model results and the completion of the model archive are required prior to Bureau approval of the information product containing the model results. Model archive implementation details are provided in the IMPLEMENTATION GUIDELINES section.

All new WMA models that are described in USGS reports or other information products (as previously defined) with publication and completion dates beginning October 1, 2015 are subject to its requirements. Groundwater models that have completion dates between October 1, 1993 and the effective date of this memorandum are subject to prior OGW model archival policies.

The archive principally is for internal USGS access and use. The release of information from the archive is subject to compliance with any existing Federal government, USGS and (or) WMA policies that may apply to the public release of such information. Unpublished information in the model archive requires review and approval prior to its release to the public as specified in SM 502.5. Care must be taken to ensure that no personally identifiable information is inadvertently released as specified by the Privacy Act in SM 319.1. Additionally, information may not be released on the specific locations of critical water infrastructure as specified in Guidance on Release of Sensitive Water Related Information.

IMPLEMENTATION GUIDELINES

The WMA model archive policy is implemented and managed by the responsible Center where the model application is developed and published. The need for a model archive should be identified in relevant project proposals and guidance on archiving project models should be incorporated into each Center's Quality Assurance Plans. Oversight of policy implementation is provided by Water Science Field Team (WSFT) Specialists. The oversight comprises two principal components, one focused on individual model applications that are reviewed before the relevant information product receives Director's approval, and one focused on the Center model archive. The status of the model archives are formally examined as a component of WMA triennial WSC technical reviews and NRP triennial reviews.

To the maximum extent possible, electronic model archives should reside in a location that is available for review at the originating Center. In some instances, an entire electronic model archive cannot physically reside at a Center. For example, if specific datasets are too large or proprietary restrictions prohibit accommodation in the Center's model archive, then these datasets can reside at another location if a persistent link or a pointer from the Center model archive to the remote location is included in the archive. In this case, the remote storage location must be permanent and readily accessible to the USGS.

The steps for model archive approval are outlined in the Model Archive Verification and Approval Form (Attachment 1), which is uploaded to the Information Product Data System (IPDS) and provided to the Bureau Approving Official (BAO) with the information product package. The model archive review is a formal requirement of the model report review process as noted in the POLICY section of this memo and must be completed prior to submittal of packages to the BAO for Director's (delegated) approval.

The steps for model archive verification and approval are:

1. The originating Center is responsible for creating and populating an electronic model archive with the appropriate directories and files. (See Attachment 2 for recommendations on electronic model archive design and implementation.) If an electronic model archive is not required, the reason that it is not required is indicated on the Model Archive Verification and Approval Form.
2. The Center Director or their designee reviews the electronic model archive to ensure compliance with this policy memorandum. The review ensures the following:
   - The contents of the model archive are complete (see Attachment 2).
   - The model contained in the archive runs to completion without errors.
   - Output generated from the archived model simulation agrees with the text, figures, and tables in the information product.
   - If an electronic model archive cannot be constructed in the recommended manner for technical reasons, the deviations and alternative approach are noted and described on the Model Archive Verification and Approval Form.
3. After the review is completed and issues encountered as part of the review are addressed, the Center Director or their designee completes the Model Archive Verification and Approval Form (Attachment 1). By signing the Model Archive Verification and Approval Form, the Center Director or designee assures that the model has been archived properly and that it is in compliance with WMA model archive policies.
4. The Center notifies the WSFT Specialist that the Model Archive Verification and Approval Form, and if required, the information product and archive are available for model archive verification. Centers submit the Model Archive Verification and Approval Form to the WSFT Specialist before the information product is submitted to the BAO for approval.
5. The WSFT Specialist ensures that the model archive, if required, and the Model Archive Verification and Approval Form have been adequately reviewed and proceeds with verification. If a model archive is required, the Specialist may:
   a. Rely heavily on the Center's review and perform a spot-check verification;
   b. Conduct a full review and verification himself/herself;
   c. Identify another qualified person to review and verify the model archive.

   If a model archive is not required, the Specialist should determine whether the justification given for not requiring an archive is sufficient and appropriate.

After any remaining issues are addressed by the author(s), the WSFT Specialist will communicate his/her verification of the Model Archive Verification and Approval Form and the model archive, if required, to the Center by signing the Model Archive Verification and Approval Form as indicated.

6. The Center Director transmits the information product, including the completed Model Archive Verification and Approval form, to the BAO for approval. The Model Archive Verification and Approval Form should be included with the information product files in IPDS.

The WMA electronic model archive does not preclude the need for investigators to adequately describe and document model applications in their information products. Although the level of documentation will vary depending on the project objectives and the complexity of the simulations, OGW Technical Memorandum 96.04 outlines the appropriate level of documentation for groundwater modeling information products. It is suggested that products documenting surface-water and water-quality models follow similar guidelines.

The Center's model archive should contain a catalog of all modeling applications contained in the archive, as well as the Center's archives maintained off-site, if necessary.

DISCIPLINE-SPECIFIC DETAILS

Because of the variety of models and other limitations, discipline-specific details related to the policy and guidelines are clarified in this section. These clarifications are not departures or exceptions from the policy described in this memorandum.

Office of Groundwater

This policy and guidance expands existing OGW model and aquifer-test archive policies in that all groundwater models—not just numerical groundwater models—potentially are subject to electronic model archiving. This includes, for example, analytical solutions to groundwater flow and transport equations and statistical models, such as regression, neural network and genetic-algorithm models for analyzing groundwater data. In some cases, no additional work will be required by this policy, such as for aquifer test analyses, which already must be archived under OGW Technical Memorandum 2009.01. To determine whether or not a model should be electronically archived, authors should carefully consider whether or not the model results can be reproduced from the report. For example, an electronic model archive is not required for simple Darcian flow calculations and linear regression equations but it is required for more complicated models or perhaps for other models that require large input data sets.

In response to previous OGW policies on electronic model archiving, most Centers have developed a complete database of recent groundwater models. In order to maintain the integrity of this resource, groundwater models archived as supplemental information in an outside publication must also be included in the Center's groundwater model archive.

The WSFT Groundwater Specialist reviews the information product to ensure that it is in compliance with USGS technical policies and includes their recommendation of approval to the BAO.

Office of Surface Water

Any models related to the streamgaging data program (indirect measurements, stage-discharge rating curves, including theoretical rating curves, index velocity regression models, estimation of AEP for observed flood peaks, etc.) are not included in this policy. These types of models must be archived in the streamgage station digital and paper files, following procedures described elsewhere (Water Science Center Quality Assurance Plans, Benson and Dalrymple, 1967; Kennedy, 1984; Davidian, 1984; Levesque and Oberg, 2012; OSW Technical Memorandum 2013.01).

The WSFT Surface Water Specialists may review the modeling report containing the results of the model application to be electronically archived at their discretion or upon request through the Chief of the WSFT.

Office of Water Quality

The WSFT Water Quality Specialists may review the modeling report containing the results of the model application to be electronically archived at their discretion or upon request through the Chief of the WSFT.

William L. Cunningham
Digitally signed by William L. Cunningham
DN: cn=William L. Cunningham, o=Office of Groundwater, ou=USGS, email=wcunning@usgs.gov, c=US
Date: 2014.12.18 14:59:46 -05'00'

William L. Cunningham
Chief, Office of Groundwater

Robert R. Mason, Jr
Digitally signed by Robert R. Mason, Jr
DN: cn=Robert R. Mason, Jr, o=Office of Surface Water, ou=USGS, email=rrmason@usgs.gov, c=US
Date: 2014.12.18 15:00:14 -05'00'

Robert R. Mason, Jr.
Deputy Chief, Office of Surface Water

Donna M. Myers
Digitally signed by Donna M. Myers
DN: cn=Donna M. Myers, o=Office of Water Quality, ou=USGS, email=dnmyers@usgs.gov, c=US
Date: 2014.12.18 15:00:47 -05'00'

Donna N. Myers
Chief, Office of Water Quality

Distribution: GS-W All

REFERENCES

Benson, M.A. and Dayrymple, Tate, 1967, General field and office procedures for indirect discharge measurements, U.S. Geological Survey Techniques and Methods Book 3, Chapter A1, 30 p.

Davidian, Jacob, 1984, Computation of water-surface profiles in open channels, U.S. Geological Survey Techniques and Methods Book 3, Chapter A15, 48 p.

Kennedy, E.J., 1984, Discharge ratings at gaging stations, U.S. Geological Survey Techniques and Methods Book 3, Chapter A10, 59 p.

Levesque, V.A. and Oberg, K.A., 2012, Computing discharge using the index velocity method, U.S. Geological Survey Techniques and Methods 3-A23, 148 p.

Office of Groundwater Technical Memorandum No. 2011.01, Groundwater Flow and Transport Model Archival, accessed December 26, 2012 at http://water.usgs.gov/admin/memo/GW/gw11.01.html

Office of Groundwater Technical Memorandum 2009.01, Update on Guidance for the Preparation, Approval, and Archiving of Aquifer-Test Results, accessed June 17, 2013 at http://water.usgs.gov/admin/memo/GW/gw09.01.html

Office of Groundwater Technical Memorandum No. 2005.02, Policy on documenting the results of new simulations using previously published ground-water models, accessed December 26, 2012 at http://water.usgs.gov/admin/memo/GW/gw05.02.html

Office of Groundwater Technical Memorandum No. 00.02, Update of the National Policy to Archive Ground Water Flow and Transport Models, accessed December 26, 2012 at http://water.usgs.gov/admin/memo/GW/gw00.02.html

Office of Groundwater Technical Memorandum No. 97.01, Policy on documenting the use of ground-water simulation in project reports, accessed December 26, 2012 at http://water.usgs.gov/admin/memo/GW/gw97.01.html

Office of Groundwater Technical Memorandum No. 96.04, Policy on documenting the use of ground-water simulation in project reports, accessed December 26, 2012 at http://water.usgs.gov/admin/memo/GW/gw96.04.html

Office of Groundwater Technical Memorandum No. 93.01, Establishment of a National Policy to Archive Ground-Water Flow and Transport Models, accessed December 26, 2012 at http://water.usgs.gov/admin/memo/GW/gw93.01.html

Office of Groundwater Branch Technical Memorandum No. 75.11, Reporting and documentation of aquifer modeling studies, accessed December 26, 2012 at http://water.usgs.gov/admin/memo/GW/gw75.11.html

Office of Management and Budget Memorandum M-13-13, Open Data Policy, Managing Data as an Asset, accessed January 6, 2014 at http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

Office of Surface Water Technical Memorandum 2013.01 dated December 5, 2012, Computation of annual exceedance probability (AEP) for characterization of observed flood peaks, accessed November 25, 2014 at http://water.usgs.gov/admin/memo/SW/sw13.01.pdf

Office of Water Information Technical Memorandum dated July 17, 2008, Guidance on Release of Sensitive Water Related Information, accessed December 13, 2013 at http://nwis.usgs.gov/communications/2008news/080717sensitive_data_guidelines.html

Survey Manual 502.2, U.S. Geological Survey Fundamental Science Practices: Definition of data documentation accessed April 1, 2013 at http://www.usgs.gov/usgs-manual/500/502-2.html

Survey Manual 502.5, U.S. Geological Survey Fundamental Science Practices: Safeguarding Unpublished U.S. Geological Survey Data, Information, and Associated Scientific Materials accessed December 13, 2013 at http://www.usgs.gov/usgs-manual/500/502-5.html

Water Resources Division Scientific Records Committee, Guidance for the Use of Scientific Records Disposition Schedule 1400: dated June 2006, accessed December 13, 2013 at http://water.usgs.gov/usgs/srm/index.html

Water Resources Discipline Scientific Records Disposition Schedules 1400-02. Research and Investigative Case Files accessed April 18, 2013 at http://www.usgs.gov/usgs-manual/schedule/432-1-s2/index.html#gw

Water Resources Discipline Policy Memorandum 2009.01. Policy on the distribution of publicly available USGS Water Resources Software on the Internet, accessed September 11, 2013 at http://water.usgs.gov/admin/memo/policy/wrdpolicy09.01.html

White House Office of Science and Technology Policy Memorandum, 22 February 2013 Increasing Access to the Results of Federally Funded Scientific Research, accessed January 6, 2014 at http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

ATTACHMENT 1

Model Archive Verification and Approval Form (include this form in IPDS prior to Bureau-level report approval)

Project Title: _____

Project Chief: _____

Model Description: _____

The model and report do not require an electronic archive:

Reason(s)

- ☐ Model results can be reproduced solely based on the contents of the USGS report or information product.
- ☐ The model archive requirement is covered by another WMA or USGS Archiving Policies.
- ☐ The results are described in an outside publication and none of the USGS coauthors were involved in development of the model application
- ☐ Other reasons for exception from policy with short description.

_____

If the electronic archive is required, complete the rest of this form; otherwise proceed to the verification and approval and signature step at the end of this form.

Electronic model archive review and verification checklist:

1) Files in archive
   Archive location (path) _____
   - ☐ Contents directory
   - ☐ Readme file or files (identifies modeling software and version)
   - ☐ modelgeoref.txt file
   - ☐ Source code and executables specific to the model application including source code and executables
   - ☐ Inputs
   - ☐ Outputs
   - ☐ Calibration targets
   - ☐ Geospatial (optional)
   - ☐ Other
   - ☐ Deviations from suggested structure with short description. (This should include location of additional model files in another archive location if applicable).
     _____

2) Model execution
    - ☐ Successfully runs to completion
    - ☐ Output matches report _____ (Identify specific output checked)
3) Other comments

Archive reviewed by: _____

Center approval by: _____

Date: _____

WSFT verification by: _____

Date: _____

ATTACHMENT 2

RECOMMENDATIONS FOR STRUCTURE OF CENTER MODEL ARCHIVES

Each Center will establish an electronic model archive. The archive must be located on media that are routinely backed up as done for other mission critical information. In most cases, Centers already may have established archives for groundwater models. The addition of archive directories for surface-water and water-quality models may follow the same structure as the previously developed groundwater model archive.

Figure 1 shows an example of a recommended model archive structure for Centers. The example includes archives for groundwater, surface-water, and water-quality models. Similar to the example of the groundwater model archive, GWMARCHIVE, the surface-water and water-quality model archives are named SWMARCHIVE and WQMARCHIVE, respectively; however, other names may be used if desired. A more detailed depiction of the sub-directory/folder structure is shown for the example of a surface-water model archive. A Report Folder, located directly below the main directory, is established for each published information product **(**report**)** containing a surface-water model analysis. Each report folder should be given a name that clearly reflects the USGS report number or outside publication identifier.

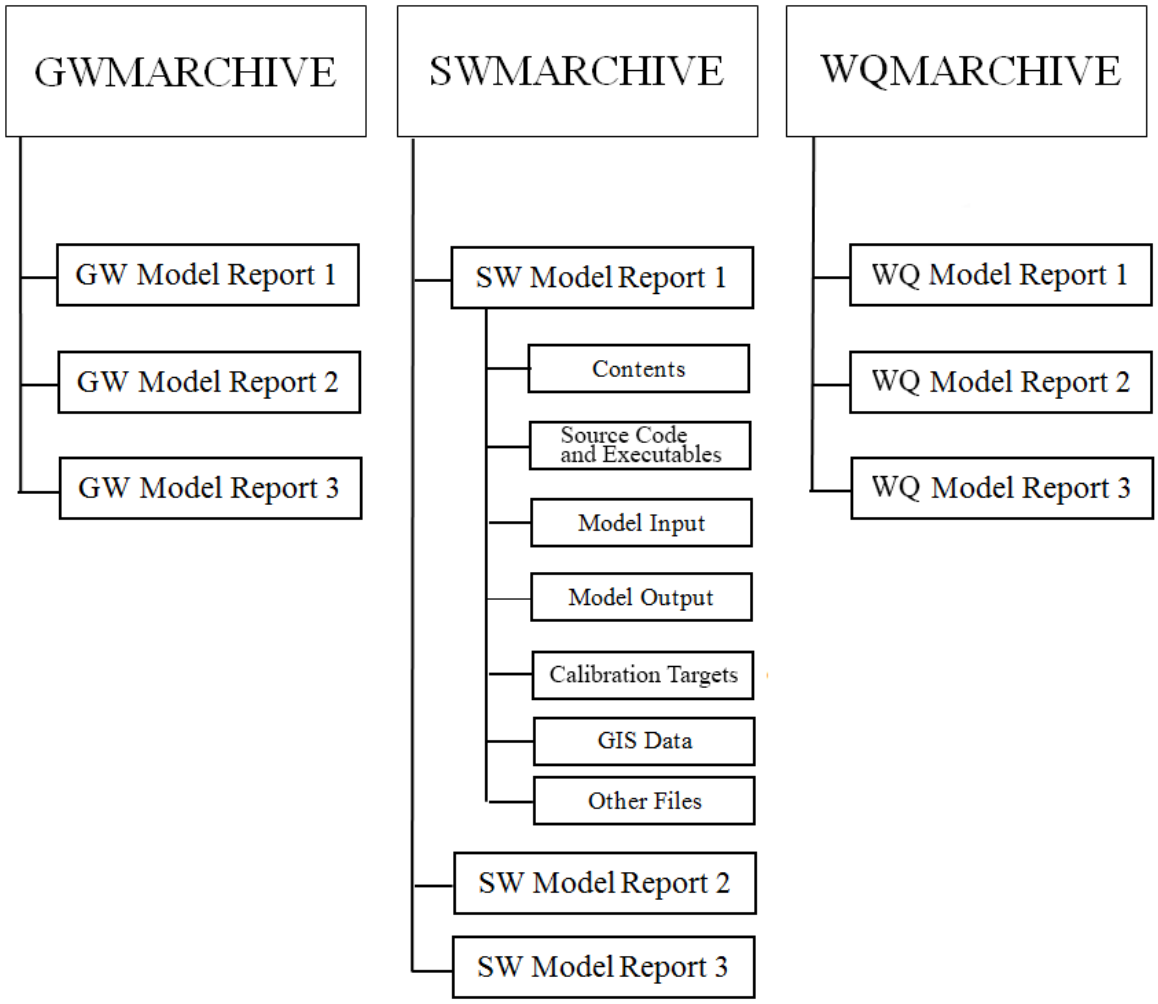

Figure 1. Example of recommended structure of a model archive showing an expanded view of the recommended directory structure under SWMARCHIVE

Another folder named CONTENTS, located within each report folder, includes a readme file or files that contain:

1. Full reference and web link for the subject report.
2. Description of the Report Archive Folder structure and of the files contained in each subfolder.
3. Identify the version of the operating system, platform, and software version used for model application
4. Text file of geo-reference information (filename: modelgeoref.txt), including latitude (or northing) and longitude (or easting) of the corners of a rectangle outlining the model study area (this information documents the area under study and allows for future map

displays of models developed by and available from the USGS). If the model is a 2-dimensional cross-section, geographic coordinates of the end points of the model should be given. If the model is a 1-dimensional model or describes processes at a single point in space, a single geographic coordinate pair can be given that describes the model 'location'.

5. Descriptions of data file formats, when appropriate.
6. Instructions for running simulation.
7. Description and location of any relevant ancillary data stored elsewhere in a permanent USGS storage system (for example, National Water Information System [NWIS] or other online archive) or persistent URL (for example a companion Data Series report).
8. If large input, output, or ancillary data are permanently stored elsewhere, then a description of where the data are stored should be included.

The archive must include the model source codes and executables (where possible), input files, macros and operating system files such as UNIX shell codes and personal computer batch (BAT) files, and model output files for each simulation described in the information product. These simulations will include (when applicable) the final calibrated results and any predictive results described in the information product. Model results of minor importance, such as interim calibration runs, should not be archived. The model output files are included to allow verification that the model reproduces the published results when the input files are rerun. Input files must be stored in the format read by the model. In cases where model input files are proprietary or machine-dependent, ASCII versions of the files also should be stored in the archive if it is possible and not overly burdensome (based on judgment of the WSFT) to generate them. The ASCII version will make it more likely that the files can be used on virtually any computer without the need for specialized or proprietary software.

The storage of ancillary data is optional, but is encouraged, especially when the data are referenced in the model information product. Examples of ancillary data that might be stored are GIS data files, data stored by pre-processor codes, or other data directly related to the model simulations. Where possible, these data should be stored in a widely supported, publically available, generic format as opposed to a proprietary format. For example ArcGIS™ offers the ability to export GIS data stored in their proprietary coverage format into formats that can be read by other applications. If ancillary data are permanently stored elsewhere, then there is no need to duplicate the storage in the model archive. In these cases, however, a description of where the ancillary data are stored elsewhere should be included in a readme file and (or) in the CONTENTS directory. Documentation of ancillary data must include a description, the source, format, and the version of software on which it was produced.

Existing models are sometimes modified to perform new simulations through recalibration with new data, extension of the model simulation period, or simulation of alternative scenarios. The new model application should be documented in a new published information product as

required by WRD Technical Memorandum 2009.01. In these cases, a model archive file should be established for the new information product, modified model, and new input and auxiliary data. If the original model and associated data were not archived, they should be archived according to policy, in order to ensure that the original model and data are preserved.

The appropriate model files must reside in the Center archive when the information product is submitted to the BAO for approval. If additional model simulations are required as a contingency for approval, these simulations should replace, or be added to the archive as appropriate.

SITUATIONS WARRANTING ADDITIONAL CONSIDERATION

There are several situations or model types for which flexibility will be needed to implement this policy. These situations or models include:

A model described in an outside publication in which one of the modelers is a USGS author: The contribution from the USGS author and modeler, regardless of the order of authorship, is subject to this policy. The contributions of authors outside USGS are not subject to this policy.

Use of a statement in the acknowledgement section of the outside publication: If the USGS author had no role in modeling, the language that follows is suggested for the acknowledgement section:

(Author's name) did not materially contribute to the model application described in this publication.

Models that use output files or data generated from other models: The files containing the output data generated from other models are included within the archive of the primary model. If the models are documented in separate information products, a cross-reference between the reports must be included in the CONTENTS directory of the archive entry for each product.

Proprietary models: If proprietary software was used for the model, an effort should be made to obtain the source code and executable files from the vendor and documentation that the software cannot be distributed must be put into the contents directory or text file. If the executable files are stored in the archive, the software version and the hardware and operating system for which the code was compiled and the executables were run should be documented.

Models with very long run times: The modeler will schedule the model archive review in advance to accommodate a long model run time. The model archive review is scheduled accordingly in advance and in anticipation of providing evidence that the modeled output and other appropriate results and files that match the information product are available in the archive at the time of the review.

Models with large input or output data files: Files can be stored outside the Center archive as long as the model archive points to a readily accessible and permanent location.

Interdisciplinary models: A "primary" discipline for which to attach the archive is selected by the model developer.

The model domain is not defined by geographic coordinates: If there is no geographic boundary, the "modelgeoref.txt" file will indicate "not applicable."

The model domain has non-rectangular boundary coordinates: Approximate a rectangle that fully encompasses the model domain and provide the appropriate geographic coordinates of the rectangle in the "modelgeoref.txt" file.

Models with a stochastic component: For model analyses with a stochastic component that require many thousands or millions of runs (such as a Monte-Carlo analysis, for example), it may not be possible to archive input and output for all of the simulations. In this case, it is important that the archive contains enough information to reproduce an ensemble of the runs that contains the same statistical properties as the ensemble described in the information product.