

for the CIWG. This purpose of the CIWG is to review and vet the policies and procedures by which the MMAP will manage its computing and data activities to ensure reliable and equitable access to computing and data resources among the partner organizations and the integrity of the data products resulting from the MMAP research program. As a minimum, the CIWG will periodically review and recommend modifications to the contents of the System Description Document. In addition the CIWG will manage the submission of computing resource proposals submitted by MMAP researchers to avoid competing submissions as well as to ensure equitable access to the awarded resources. Membership in the CIWG will be open to all members of the MMAP community.

1.2 System Description Document

2 Funded Personnel

Staffing for the Data Management function of MMAP is provided by UCSD San Diego Supercomputer Center (SDSC) personnel. The team is led by Dr. John Helly, Director of the Laboratory for Earth and Environmental Science (LEES) and Associate Director for Computing for the Science and Technology Center. The other partially funded team members include Dr. Don Sutton and John Weatherford; both programming staff within LEES and full-time staff at the SDSC. There is also a full-time position for a graduate student beginning in the second year and continuing for the remaining four years. All the resources of the SDSC are also available for consultation and assistance as would be the usual case for staff members and for allocated researcher.

3 Computing Allocations

One of the primary tasks of the MMAP is to obtain and manage large computing allocations from the HPCC resources available to United States researchers. As stated above, the CIWG will be the focal point for vetting proposals that go forward to the various program opportunities to ensure high-quality proposals compliant with the requirements of the various resource providers. The Associate Director for Computing will develop a schedule of the proposal opportunities and publicize these to the MMAP community to encourage maximum participation and awareness of the resource opportunities.

3.1 NSF

3.1.1 SDSC

Immediate access to startup and testing allocations are available through the Teragrid and we will immediately begin developing and submitting allocation proposals in the normal proposal cycles of the Teragrid and Cyberinfrastructure Partnership (CIP) as well as the national laboratories. Table 1 shows the resource allocation matrix for SDSC. It describes three types of allocations and the proposal opportunities for 2006. This process will evolve as the NSF cyberinfrastructure evolves but something equivalent to it will be in effect in subsequent years.

Table 0 SDSC Resource Allocation Categories and Schedule

5/18/06 8:43 AM

Deleted: 1

	Service Unit (SU) limit	Proposals Accepted	Start Dates
Development Allocations Committee (DAC)	0-10,000 (0-30,000 for TeraGrid)	Proposals accepted at any time and reviewed as they are submitted.	Immediately following successful review.
Medium Resource Allocations Committee (MRAC)	10,001-200,000 (30,001-200,000 for TeraGrid)	<ul style="list-style-type: none"> ● Dec.16,2005-Jan.13,2006 ● Mar.17-Apr.14,2006 ● Jun.16-Jul.14,2006 ● Sep.15-Oct.13,2006 	Approx. two months following submission, on Apr. 1, July 1, Oct. 1 and Jan. 1.
Large Resource Allocations Committee (LRAC)	200,001 and up	<ul style="list-style-type: none"> ● Dec.16,2005-Jan.13,2006 ● Jun.16-Jul.14,2006 	Approx. two months after submission, on Apr. 1 and Oct. 1.

3.1.2 NCAR

The procedure at NCAR is documented at <http://www.cisl.ucar.edu/main/computers.html>

3.2 DOE

3.2.1 LLNL

The procedure at LLNL is documented at https://www.llnl.gov/lcforms/policies/onsite_access_policy.html.

3.2.2 ANL

3.3 NASA

3.3.1 LaRC

3.3.2 GSFC

3.3.3 Ames

4 Data Access Policy

The MMAP data policy includes three specific sections designed to express shared network policies regarding the release of MMAP data products, user registration for accessing data, and the licensing agreements specifying the conditions for License.

4.1 Timely Release of Data

Data and information derived from publicly funded research at the MMAP, totally or partially from MMAP funds from NSF, Institutional Cost-Share, or Partner Agency or Institution where a formal memorandum of understanding with MMAP has been established, are made available online with as few restrictions as possible, on a nondiscriminatory basis. MMAP scientists should make every effort to release data in a timely fashion and with attention to accurate and complete metadata.

4.2 Classification of Data

There are two data types:

4.2.1 Type I Data

Data are to be released to the general public according to the terms of the *General Public License Agreement* within 2 years from collection and no later than the publication of the main findings from the Data and,

4.2.2 Type II Data

Data are to be released to restricted audiences according to terms specified by the owners of the data. Type II data are considered to be exceptional and should be rare in occurrence. The justification for exceptions must be well documented and approved by the *Working Group*.

While the spirit of this document is to promote maximum availability for data in either Type I or II status, there are criteria by which priority for data release may be determined. Primary observations collected for core research activities directly supported by MMAP research must receive the highest priority for data release. Data collected by other sources to which MMAP supported research has added value is also a high priority subject to intellectual property rights of the data owners. Other types of data including non-MMAP data that acquired for MMAP research, student thesis data, or legacy data that already suffer from inadequate documentation or format obsolescence may be ranked a lower priority by a site with justifications provided in their data management policy. Finally, some data may be determined of lowest priority for archiving on the grounds that they are interim data that led to final products that carry the scientific value. These might include data files created during stages within an analytic workflow, raw or replicate data values that were subsequently aggregated or processed for release.

4.3 Requirements of the User

The access to all MMAP data is subject to requirements set forth by this policy document to enable data providers to track usage, evaluate its impact in the community, and confirm users' acceptance of the terms of acceptable use. These requirements are standardized within MMAP to provide contractual exchange of data between data providers and Users. To effect this relationship, the following information may be required directly or by proxy prior to the transference of any data object:

4.3.1 Registration

1. Name
2. Affiliation
3. Email Address
4. Phone Number
5. Acceptance of the General Public License Agreement or Restricted License, as applicable.

Data providers wishing to impose further requirements beyond these are encouraged to include them in their Restricted License Agreements accompanying the data.

4.4 License Agreements

Data released by MMAP sites or the network will be accompanied with a use agreement that specifies the conditions for License. For Type I data, this shall be the General Public License Agreement (see appendix II). This document specifies general roles and the obligations and rights enjoyed by each regarding the use of most Data released for general public use. For Type II Data, a Restricted License Agreement must be provided with the Data that identifies the specific restrictions on the use of the data and their justification. Because these are expected to be unique to the Data, no template is provided although in most cases the General Public License Agreement can be modified to serve. Grounds for restricting data may include the need to restrict access to species, habitats or cultural resources protected by legislation; rights of privacy granted by human subjects legislation; or protection of intellectual, financial or legal rights over the data held by a third party.

This policy becomes effective when approved by the MMAP Network Coordinating Committee. It may be revised by, or at the request of, the same body.

4.4.1 General Public License Agreement

4.4.1.1 Definitions

- *Data* – Digital data and its metadata derived from any research activity such as field observations, collections, laboratory analysis, experiments, or the post-processing of existing data and identified by a unique identifier issued by a recognized cataloging authority such as a site, university, agency, or other organization.
- *User* - individual to whom access has been granted to this data, including his or her immediate collaboration sphere, defined here as the institutions, partners, students and staff with whom the User collaborates, and with whom access must be granted, in order to fulfill the User's intended use of the data.
- *Creator* - individual or institution that produced the data
- *Owner* – individual or institution that holds intellectual property rights to the data. Note that this may or may not be defined as a legal copyright. If no other party is designated in the metadata as Owner, it may be presumed that these rights are held by the Creator.
- *Distributor* - individual or institution providing access to the data.
- *Contact* - party designated in the accompanying metadata of the data as the primary contact for the data.

4.4.1.2 Conditions of Use

The re-use of scientific data has the potential to greatly increase communication, collaboration and synthesis within and among disciplines, and thus is fostered, supported and encouraged. Permission to use this Data is granted to the User free of charge subject to the following terms:

- 1) Acceptable use. Use of the data will be restricted to academic, research, educational, government, recreational, or other not-for-profit professional purposes. A user is permitted to produce and distribute derived works from this data provided that they are released under the same license terms as those accompanying this data. Any other uses for the data or its derived products will require explicit permission from the data owner.

2) Redistribution. The data are provided for use by the user. The metadata and this license must accompany all copies made and be available to all users of this data. The User will not redistribute the original data beyond this collaboration sphere.

3) Citation. It is considered a matter of professional ethics to acknowledge the work of other scientists. Thus, the User will properly cite the Data in any publications or in the metadata of any derived data products that were produced using the Data. Citation should take the following general form: Creator, Year of Data Publication, Title of Data, Publisher, Data identifier.

4) Acknowledgement. The User should acknowledge any institutional support or specific funding awards referenced in the metadata accompanying this Data in any publications where the Data contributed significantly to its content. Acknowledgements should identify the supporting party, the party that received the support, and any identifying information such as grant numbers.

5) Notification. The User will notify the Data Contact when any derivative work or publication based on or derived from the Data is distributed. The User will provide the data contact with two reprints of any publications resulting from use of the Data and will provide copies, or on-line access to, any derived digital products. Notification will include an explanation of how the Data was used to produce the derived work.

6) Collaboration. The Data has been released in the spirit of open scientific collaboration. Users are thus strongly encouraged to consider consultation, collaboration and/or co-authorship with the Data Creator.

By accepting this data, the user agrees to abide by the terms of this agreement. The data owner shall have the right to terminate this agreement immediately by written notice upon the user's breach of, or non-compliance with, any of its terms. The user may be held responsible for any misuse that is caused or encouraged by the user's failure to abide by the terms of this agreement.

4.4.1.3 Disclaimer

While substantial efforts are made to ensure the accuracy of data and documentation contained in this data, complete accuracy of data and metadata cannot be guaranteed. All data and metadata are made available "as is". The user holds all parties involved in the production or distribution of the data harmless for damages resulting from its use or interpretation.

5 CyberInfrastructure

The cyberinfrastructure for the MMAP will evolve but a preliminary configuration is depicted in the diagram below. This architecture integrates national resources for the benefit of the MMAP research program and exhibits challenges in managing the resources across different federal organizations and national centers. This will be a key management function.

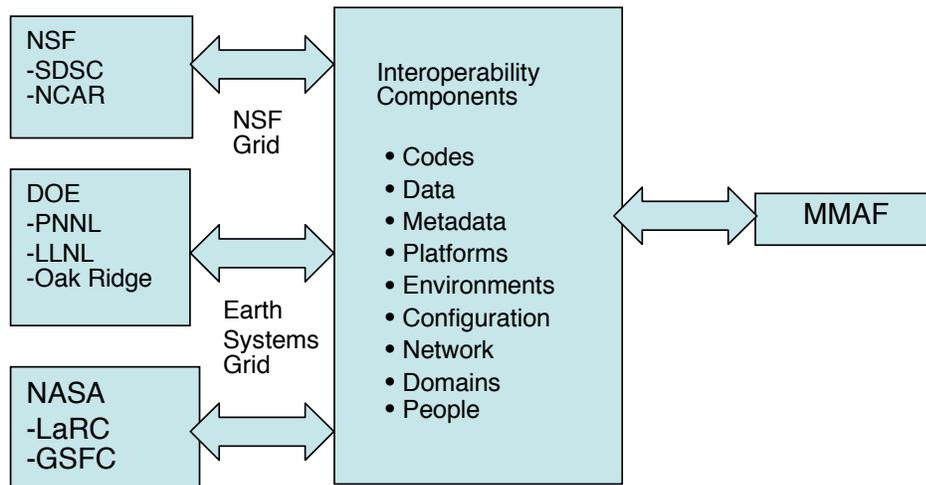


Figure 1 Cyberinfrastructure emphasizing external interfaces.

At the NSF and other national computing resources evolve, MMAP will continue to compete for access to them through the usual allocation proposal processes.

5.1 Interoperability

As indicated in the figure above, interoperability will be a key feature of the MMAP cyberinfrastructure. We will support efforts to interoperate with the Earth Systems Grid, existing community resources such as those provided and supported by NCAR and develop and maintain the necessary interface descriptions to enable effective use of the supported products and computing resources.

5.1.1 External Interfaces

5.1.2 Internal Interfaces

5.2 Education, Outreach and Training

5.2.1 GoogleEarth KML for MMAP Activities

5.3 Metadata

- Metadata documenting data of any type will be made available when, or before, the Data itself is released according to the terms above.
- All metadata will be publicly available regardless of any restrictions on access to the data.

All metadata will follow MMAP standards and conventions and will minimally contain adequate information for proper citation, access, contact information, and discovery and reuse.

5.4 Data Resources

5.4.1 SDSC

SDSC has committed 15 TB of spinning disk as part of its match and the CIWG will prepare a data allocation request for additional storage resources in the hybrid tape-disk SAM-QFS system.

5.5 Supported Products

Supported products will include codes, documentation and data that will be maintained by the MMAP for the benefit of the community.

5.5.1 Version Control System

5.6 Accession Methods and Web-Presence

Accession to supported products will be through the Internet using appropriate methods for the task. There will be a central web-portal to provide 'one-stop-shopping' through the resources of the MMAP and this site will integrate and re-direct the user to appropriate secondary servers for access to supported products.

5.6.1 Content Management System (e.g., Plone)

This is a collaborative environment for self-publishing of digital content that can be managed without a 'webmaster' as a bottle-neck.

5.6.2 Digital Library Server

This is a server for providing for search and retrieval of supported products and related data published under the auspices of MMAP.

5.6.3 Web-portal (<http://mmap.org> (???)

5.7 Quality Control Procedures

Quality control of data and software will be the responsibility of the developers and the science team using them with review by the CIWG for compliance with MMAP standards and conventions. MMAP will develop procedures for assessing data and software quality commensurate with current best practices in the computer science and atmospheric research communities.

5.8 Maintenance

Maintenance of the cyber-resources is the responsibility of the home institutions housing the resource.

5.9 Backup and Recovery

Backup and recovery will be accomplished using the HPSS storage system at SDSC as well as the dedicated disk resource of 15TB committed to the MMAP project. Off-site storage of data, including software, will be stored in the MMAP facilities in Fort Collins on an appropriate medium. Backups will be run on a regular basis on a schedule that will be developed as the project evolves. It will be documented in the SDD.

