**The OCC Mission**

The Open Commons Consortium (OCC):

1. Manage storage and data commons infrastructure, such as the Open Science Data Cloud Public Data Commons, the NOAA Data Commons, and the Biomedical Data Commons. Provide index, metadata, transfer and other value services supporting data commons activities.
2. Manage cloud computing infrastructure such as the Open Science Data Cloud, to support scientific, environmental, medical and health care research.
3. Manage cloud computing testbeds, such as the Open Cloud Testbed, to improve cloud computing software and services.
4. Develop reference implementations, benchmarks and standards, such as the MalStone Benchmark, to improve the state of the art of cloud computing.

MalStone that is specifically designed to measure the performance of cloud computing middleware that supports the type of data intensive computing common when building data mining models. We also introduce MalGen, which is a utility for generating data on clouds that can be used with MalStone.

1. Sponsor workshops and other events related to cloud computing and data commons to educate the community.

*The Open Commons Consortium is a consortium managed by the Center for Computational Science Research, Inc., which is an Illinois based not-for-profit corporation.*

# Working Groups

Working groups support the interest and activities of OCC Members. The current working groups include:

### The Open Science Data Cloud (OSDC) Working Group

This is a working group that manages and operates the Open Science Data Cloud (OSDC), which is is a petabyte scale science cloud for researchers to manage, analyze and share their data. Individual researchers may apply for accounts to analyze data hosted by the OSDC. Research projects with TB-scale datasets are encourage to join the OSDC and contribute towards its infrastructure.

The Institute for Genomics and Systems Biology at the University of Chicago uses the OSDC as the basis for Bionimbus, a cloud for genomics and related data. Partners of the OSCD include: the [Moore Foundation](http://www.moore.org/), which is providing the hardware infrastructure for Phase 1 of the OSDC (2011-2014); Yahoo, which contributed the equipment for the Proof of Concept OSDC (2008-2010); and Cisco, which is providing the Cisco C-Wave to connect the several geographically distributed OSDC data centers. Funding to connect international partners to the OSDC is provided by the [NSF](http://www.nsf.gov/news/news_summ.jsp?cntn_id=118681).

You can apply for access to the Open Science Data Cloud by applying for a resource allocation [here](http://www.opensciencedatacloud.org/apply/).

Currently, anyone can freely access data from the [OSDC Public Data Commons](https://www.opensciencedatacloud.org/publicdata/) but to receive a resource allocation with compute, you generally need to be associated with one of the research projects of an OCC Member.

### NOAA Data Alliance Working Group

The OCC National Oceanographic and Atmospheric Administration (NOAA) Data Alliance Working Group supports and manages the NOAA data commons and the surrounding community interested in the open redistribution of NOAA datasets.

Working group activities include: i) developing and operating a petabyte-scale NOAA data commons; ii) working with the alliance community to determine key datasets to be stored in the NOAA data commons; iii) working with the alliance community to prioritize key middleware and data services; iv) implementing functional networking and connectivity between NOAA and the OCC NOAA data commons; v) working with other NOAA data alliances to adopt data peering and standards for metadata and ID services; and vi) developing a self-sufficient and sustainable ecosystem through which NOAA can reliably release key datasets to the public.

### Project Matsu Working Group

Project Matsu is a collaboration between the NASA Goddard Space Flight Center and the Open Commons Consortium to develop open source technology for cloud-based processing of satellite imagery to support the earth science research community as well as human assisted disaster relief. This working group develops and operates the OCC [Matsu Cloud](http://matsu.opensciencedatacloud.org). Matsu resources are also available to those who needed elastic computing capacity to assist with flood and fire disaster relief.

### Biomedical Data Commons (BDC) Working Group

The [Biomedical Data Commons (BDC)](http://cdis.uchicago.edu/resources/bcc/) is cloud-based infrastructure that provides secure, compliant cloud services for managing and analyzing genomic data, electronic medical records (EMR), medical images, and other PHI data. It provides resources to researchers so that they can more easily make discoveries from large complex controlled access datasets. The BDC provides resources to those institutions in the BDC Working Group. It is an example of what is sometimes called condominium model of sharing research infrastructure in which the research infrastructure is operated by a consortium of educational and research organizations and provides resources to the consortium.

Working group activities include: i) developing and operating a secure, compliant cloud for human genomic data; ii) developing and operating a secure cloud-based enclave that can be used for storing and analyzing electronic medical and health records in a HIPAA-compliant environment; and iii) contributing towards a petabyte-scale repository of open and controlled access biomedical data.

### The OCC Testbed Working Group

This working group manages and operates the OCC Testbed. The OCC Testbed (OCC-T) is a geographically distributed cloud testbed spanning four data centers and connected with 10G and 100G network connections. The OCC-T is used to develop new cloud computing software and infrastructure. Membership in this working group is limited to OCC members who contribute computing, networking, or other resources to the OCC Testbed. A current focus of the OCC-T is developing an OpenFlow enabled version of Hadoop, a project support by the NSF.