

Data Intensive Cyber Environments Center

Reagan Moore, Director

The Data Intensive Cyber Environments (DICE) Center at Carolina is an interdisciplinary community of scholars and science and technology experts who have come together to address two overarching, interrelated challenges in today's digital world:

- The need to share and collaborate with growing digital data collections to solve urgent problems in science, medicine, and other areas for North Carolina and beyond.
- The need to preserve our North Carolina and national digital history long-term, to ensure society's "memory" is accessible to future generations.

To make this possible, the DICE Center is building on internationally recognized experience and software cyberinfrastructure to accomplish the following:

- Develop an interdisciplinary model for a campuswide data-management cyberinfrastructure.
- Collaborate with groups across the UNC system on issues of distributed data sharing and management, digital preservation, data intensive computing, digital library systems, and data curation.
- Sustain national and international leadership in innovating new approaches for sharing and integrating data through federation or interconnection of previously stand-alone repositories.

Major Accomplishments During the Past Year

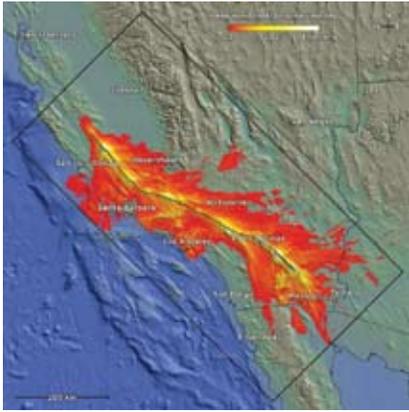
The DICE Center is collaborating with multiple groups at Carolina to link data repositories across the UNC system and the state to make them accessible and easy to use for researchers, businesses, and the government. Highlights include the following:

- The DICE Center released version 2.4.1 of the iRODS integrated Rule-Oriented Data Grid, which provides advanced features including support for extensible metadata tables, enforcing storage quotas, and soft-links between data grids.
- DICE Center faculty published a textbook on policy-based data management, *iRODS Primer: Integrated Rule-Oriented Data System*. It is available through the Synthesis Lectures on Information Concepts, Retrieval, and Services at more than 300 universities.
- DICE Center faculty created a new class on policy-based data management at the School of Information and Library Science (SILS) for the 2010 spring semester, based on the *iRODS Primer* textbook.
- More than 10 DICE Center iRODS-based data grids and preservation environments are being installed at Carolina.
- The UNC-based Carolina Digital Repository went into production April 4, 2010, with a preservation environment based on iRODS. And a TUCASI Infrastructure Project data grid was established between NC State University and the Renaissance Computing Institute (RENCI).



Climate data gathered by ocean-crossing underwater gliders will flow to the National Climatic Data Center repository in North Carolina. This is possible thanks to "federation," connecting formerly isolated repositories using the DICE Center's advanced software iRODS. Image: RUCOOL.

- The DICE Center iRODS open source software is gaining wider use. One thousand different projects downloaded the software this year, and the rate of software downloads is increasing with each successive version.
- The DICE Center began a collaboration with SILS to develop a LifeTime Learning Digital Library for students. This will serve as a "personal digital library" or data archive for digital material acquired and created by students during their education. This approach enables students to organize, manage, and preserve a reference collection to use in their professional careers. The initial goal is to provide up to a terabyte of storage for each student, and maintain the collection for five years. This will form a "digital bridge" back to UNC for students after graduation, and will allow them to keep using and updating their collections in continuing education. This initiative can serve as the foundation for a full education cyberinfrastructure at Carolina.
- The DICE Center submitted a full National Science Foundation (NSF) DataNet proposal for the DataNet Federation Consortium. This project, if funded, will begin the formation of national data infrastructure, and the DICE Center and RENCI will serve as a hub for federation, or sharing of data, between data grids. At Carolina, the proposal also involves Information Technology Services, the Carolina Digital Repository, SILS, the Odum Institute, and RENCI.
- As a key step toward enabling researchers to share data between currently stand-alone digital repositories, the DICE Center data grid at RENCI was connected, or federated, with a data grid at the National Climatic Data Center, the Ocean Observatories Data Grid, and a CUAHSI data grid, all of which are based on DICE Center iRODS infrastructure.
- DICE Center faculty and researchers published or submitted more than 30 papers and gave more than 20 presentations to broaden impact and collaborations.



In one of the largest earthquake computations ever, researchers at the Southern California Earthquake Center simulated a Magnitude 8 earthquake. The simulations generate hundreds of thousands of gigabytes of data. These massive data collections are managed using the DICE Center-developed software iRODS. Image: SCEC.

Goals for the Coming Academic Year

- The DICE Center will implement, if awarded, the DataNet Federation Consortium Proposal. The proposal addresses all phases of the data life cycle, sustainability of reference collections, and use of research data within classrooms. The project includes participation by Duke, the University of South Carolina, Drexel University, the University of Arizona, Arizona State University, the National Climatic Data Center, and the University of California, San Diego.
- The DICE Center will work with the TUCASI Infrastructure Project to make the important digital resource of LiDAR topography data available to NC State, Duke, and Carolina, as well as orthophotographic images for the entire State of North Carolina.
- The DICE Center will seek collaborations with Data Direct Network on integration of cloud storage with institutional repositories through the iRODS data grid, broadening access to key infrastructure services.
- The DICE team will continue to release new versions of iRODS, adding advanced features required by the National Climatic Data Center, Institute of the Environment, NCB-Prepared, and multiple other projects at Carolina and in the state.
- The DICE Center will continue collaborations with national and international partners on research and development of data intensive cyber environments. This includes projects in the United Kingdom (Sustaining Heritage Access through Multivalent ArchiviNg), France (IN2P3, French National Library), Australia (Australian Research Collaboration Service), Japan (KeK), and elsewhere. National projects include the NSF Ocean Observatories Initiative and the NSF Temporal Dynamics of Learning Center.
- The DICE Center will continue to expand collaborations and provide help with the data-management needs of the Carolina community and communities across North Carolina.
- The DICE Center will teach classes on data grids in SILS and provide other educational opportunities.
- The DICE Center will support the implementation of the LifeTime Learning Library initiative in SILS and broaden participation in this initiative by other Carolina centers.