Industry Partnerships 2014

RENCI Models For Engaging with Industry

Stanley C. Ahalt, PhD
Director, Renaissance Computing Institute (RENCI)
Professor, Department of Computer Science, UNC Chapel Hill
Director, Biomedical Informatics Service, UNC-CH SOM

renci

RESEARCH \ ENGAGEMENT \ INNOVATION

RENCI Mission

Lead in R&D in advanced **cyberinfrastructure** (CI) research and development

Focus on data: **Data to Decisions**

Be an essential CI partner for:

| Triangle university research teams | Research Triangle area industries | State of NC and federal agencies |

FOCUS: complicated multidisciplinary problems and research

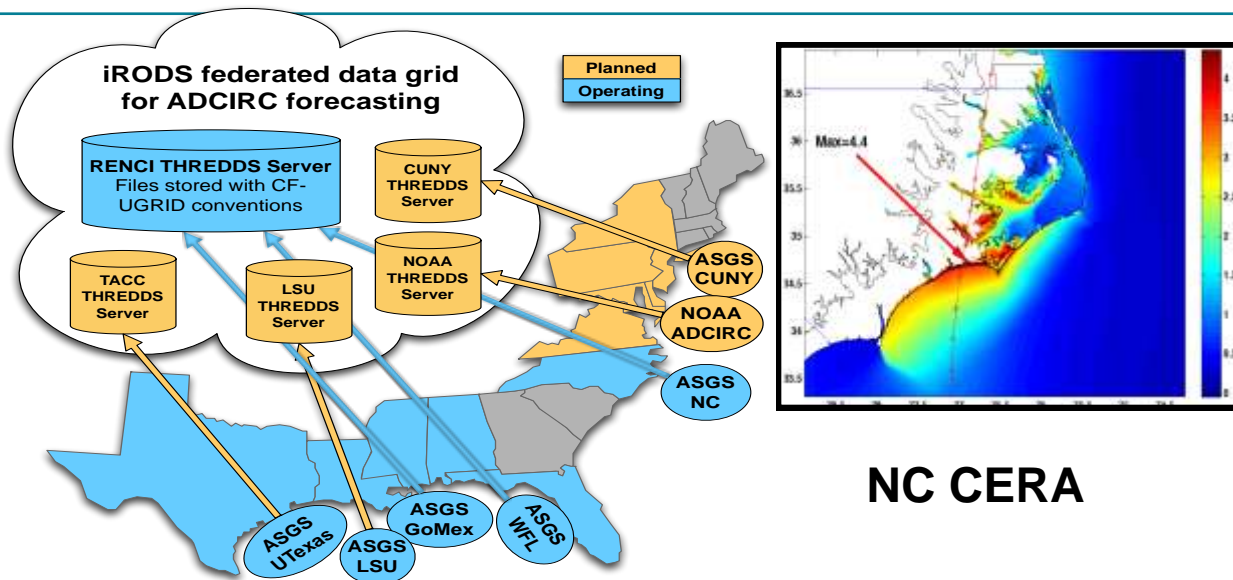renci RESEARCH \ ENGAGEMENT \ INNOVATION

RENCI Snapshot

- **Physical and Human:** Compute cluster (**~30** teraflops peak performance); **2.7** petabytes spinning disk; experimental network connecting UNC, Duke and NCSU; experts in software, bioinformatics, HPC, geoviz and domain sciences

- **Funded Data Projects**: **9** funded data projects, total funding of **$10.9M**

- **Medical Informatics:** Data tech for genomics/medical research: HPC cluster, large disks for local computations, relational DB management system, software (iRODS, Secure Medical Workspace, Hadoop, Pegasus); decision dashboards and analytics for clinical decision making.

- **Environmental Science**: Extreme event mitigation, flood plain analysis, water science software, close collaboration with Red Hat on open innovation.

- **Data Science Leadership**: **iRODS** is a leading data management framework for research; **National Consortium for Data Science** launched to put NC in Big Data leadership position.

- **Cyberinfrastructure Leadership**: *ExoGENI*: building next-gen Internet for research in Big Data era; *CI-BER*: CI for billions of EMRs; *Secure Research Workspace:* secure data handling; *Water Science Software Institute*: transforming software needed for Big Data analysis; *REACH-NC:* searchable public database of NC expertise (LinkedIn for researchers)

renci
RESEARCH \ ENGAGEMENT \ INNOVATION

What does RENCI do?

- Accelerates research via CI.
- Research development: streamline team formation and proposal creation
    - Outsources where possible: temps, free-lancers, contract work.
- Software development.
    - iRODS
    - ADCIRC+
    - Secure Research Workspace, etc.
- Almost no "production" (batch) HPC.
- Speculative investments:
    - Individual researchers
    - Emerging themes
    - Critical tools.

renci
RESEARCH \ ENGAGEMENT \ INNOVATION

Storm Surge Forecasting (ADCIRC)

iRODS federated data grid for ADCIRC forecasting

Planned / Operating

RENCI THREDDS Server — Files stored with CF-UGRID conventions

CUNY THREDDS Server

TACC THREDDS Server

LSU THREDDS Server

NOAA THREDDS Server

ASGS CUNY

NOAA ADCIRC

ASGS NC

ASGS UTexas

ASGS LSU

ASGS GoMex

ASGS WFL

NC CERA

Sandy (2012) and Irene (2011) flooding forecasts used by

- **National Hurricane Center in Miami**
- **US Coast Guard Atlantic Command**
- **Regional National Weather Service Offices**
- **State and local emergency managers**

- System uses **US NSF/NARA** funded iRODS**, NOAA NOS** gauge data, **USGS** data, **US DHS/FEMA** collected high-water mark, meteorological forecasts from **NOAA's NCEP** and **NHC**

- Very large pre-existing datasets; provides early guidance information, available about 10 minutes after official NHC forecast storm advisory

- **US DHS-funded research activity** through the DHS Coastal Hazards Center of Excellence at the University of North Carolina at Chapel Hill

- **Winner, DHS Science & Technology Impact Award, 2012**

renci — RESEARCH ENGAGEMENT INNOVATION

NCGENES – iRODS for clinical genomics

Today:

- NIH prototype to evaluate the *ethical* and *social* challenges of genomic sequencing in clinical care
- Big Data to clinically-relevant knowledge ('Clinical bins')
- Over 100 patients in the system today…

Tomorrow:

- 100M+ genomes scattered throughout the health care system
- We face a multitude of data challenges before we realize the potential of genomics in healthcare…

North Carolina Genetic Evaluation by Next-generation Exome Sequencing

Criteria:	Loci with Clinical Utility	Loci with Clinical Validity			Loci with Unknown Clinical Implications	Loci with important reproductive implications
Bins:	**Bin 1** Genes, which when mutated, result in high risk of clinically actionable condition	**Bin 2A** Low risk incidental information	**Bin 2B** Medium risk incidental information	**Bin 2C** High risk incidental information	**Bin 3** All other loci	**Bin R** Carrier status for severe AR disease

COSMOPOLITAN Relationships Cel

MEDICAL MYSTERIES

For 30 Years I Couldn't Walk,

Until I Took One Pill
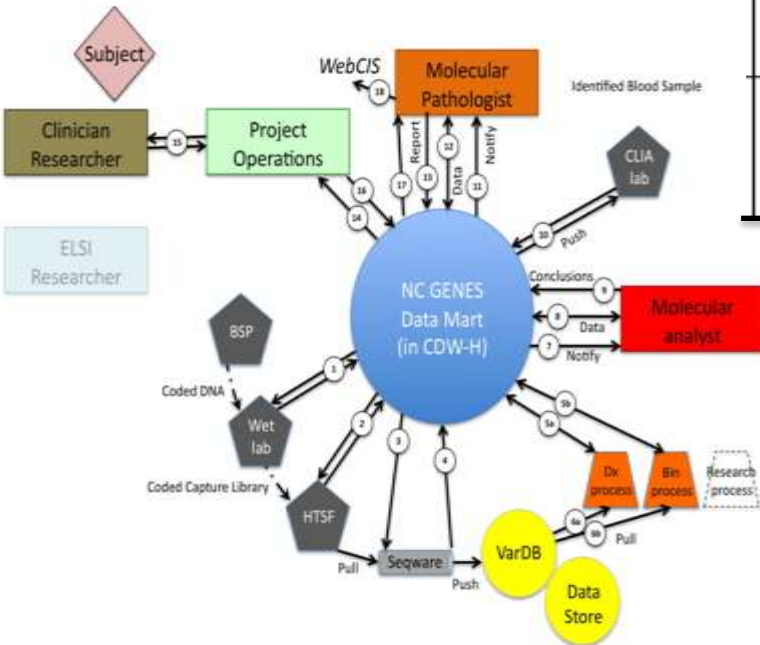
Industry-University Partnerships

The Big Picture

Univesities increasingly don't want to be isloated:

- (some) University research teams need to work with industry to make tech/knowledge transfer real AND to increase economic impact AND to understand real-world problems.

- Industry needs university researchers to tackle vital questions too risky for the private sector (profit not likely) AND to access a skilled workforce AND to gain early insights into new research outcomes.

Different strokes for different folks:

- Partnership models depend on the players, the goals, the politics, etc. **No one model will fit every need.**

renci RESEARCH \ ENGAGEMENT \ INNOVATION

The UNC/Triangle "solution:" Consortia

Time

Consortium can plant a stake in the ground quickly: significant funding and full-time staff not essential to launch.

Participation

Ability to have your voice heard, define the issues to be tackled.

Flexibility

Able to try different models, different key projects, different core foci to see what works best and to respond to changing and varied needs and interests.

Community

Consortium is a community. Building a community can become the foundation for a center (a physical place).

renci RESEARCH \ ENGAGEMENT \ INNOVATION

Example 1: iRODS Consortium

iRODS

- **Open source data management middleware tool**
- **Hundreds of installations worldwide: NASA, NOAA/NCDC, French National Library, Broad Institute**

The Issue

As iRODS transitions off NSF funding that supported a decade of development, a new sustainability model is needed.

The Approach

Build a community of iRODS use that will help sustain the software and attract new users/community members.

The Vision

Establish iRODS as the cutting-edge, **enterprise quality middleware sustained, supported and continually approved by a vibrant community of users**.

Why an iRODS Consortium?

Maintains iRODS Open Source Culture	Encourages Community Involvement	Stakeholders Are Involved In Decision Making	Revenue Flexibility: *Membership, Service, Support*	Maintains iRODS University Connection

Goals and Objectives

Sustain the development of the iRODS technology

Increase the adoption of iRODS

Develop strategic relationships around the use of iRODS

Further RENCI's work in data science: fits our expertise and is core to our mission

renci RESEARCH \ ENGAGEMENT \ INNOVATION

iRODS Consortium Membership

Current Members

- RENCI
- DICE (Data Intensive Cyber Environments Group, UNC-Chapel Hill
- Max Planck Society (Germany)

- Data Direct Networks (U.S.)
- Wellcome Trust Sanger Institute (UK)
- EMC (U.S.)

Membership Levels:
General, Sustaining, Professional

Membership Benefits (varies with level)

- Guaranteed support, consulting, training (prioritized by level)
- Committee membership (planning, working group, executive)
- Voting rights on software releases (prioritized by level)
- Sponsored events
- Recognition on Consortium materials and websites
- Access to Consortium lists

Accomplishments-Year 1

- Development of business model, membership structure, charter and bylaws

- Initial members secured at all three membership levels

- Executive director hired

- Support model established

- **Major 4.0 iRODS release on track for Spring 2014**

Released!

Marketing Plan in Development

5-Year Vision

Technology Part of the Worldwide Cyber-Infrastructure	Installable and Usable by Non-Specialists	Consortium Cash-Flow Positive	Technology Well Maintained by Consortium and Community

Example 2: National Consortium for Data Science

The Issue

Ability to manage and use data is not keeping up with its growth and complexity; solutions are targeted, rather than holistic.

The Approach

Bring together **data creators and users** in **industry** & **government** with data researchers and infrastructure developers in **academia** to discuss common challenges and find solutions.

The Vision

Develop the *science of data*, speed the transition of ideas and prototypes from the lab to the marketplace, educate next generation of data scientists.

The NCDS at a Glance

Mission

Strategize and leaders efforts to develop data science and educate future data scientists so that the power of data can be used to solve problems and drive economic growth.

Vision

Focused multi-sector, multidisciplinary data science community to solve big data challenges and drive the field forward.

Goals

- **Engage** broad communities of data experts
- **Coordinate** data science research priorities that span disciplines and industries
- **Facilitate** development education & training programs
- **Support** development of technical, ethical & policy standards
- **Apply** NCDS expertise to data challenges in science, business and government

ncds is a strategic approach to data science and big data opportunities

NCDS Components

Data Observatory

Shared, distributed infrastructure housing large organized research data; platform for data science education

Data Laboratory

R&D into critical tools and techniques for data science

Data Fellows Program

Seed grants for faculty and post-docs to work on consortium-approved projects; NCDS review panel will evaluate proposals

Industry internships for graduate students

Visiting industry data scientists at member universities

Data Science Events

Leadership Summits (Spring)

Outreach events and speakers (Fall and Spring)

NCDS Founding Members

Value Proposition

Industry Value:

Collaborative access to expertise and best practices in managing and exploiting data across academia, industry and government.

Academic Value:

Research advantages and data science leadership through collaboration with data science experts and collaborative access to data expertise across academia, industry and government

Government Value:

Mission impact, enhanced reputation, wise use of taxpayer dollars through collaborative access to data expertise across academia, industry and government

Public/Private Partnership ON Neutral Territory Territory

Strategic marketing consultants: CMG Partners CMG Partners

renci
RESEARCH \ ENGAGEMENT \ INNOVATION

Accomplishments-Year 1

Organizational:

Bylaws passed, steering committee and working groups (Data Fellows, Data Observatory/Lab, Outreach) established, n paid memberships so far.

Programmatic:

NCDS Leadership Summit *(April 2013 – kickoff event attended by Dr. Eric Green, director, NHGRI, and US Rep. David Price);* Five Faculty Fellows named *(October 2014)*; Student-Industry-Faculty career awareness event held *(April 2014)*; Observatory data sets growing, evolving.

Upcoming:

NCDS Data Innovation Showcase (May 21, 2014), Data Matters Summer Workshop Series June 23 - 28, 2014), lunchtime webinars to launch soon.

Conclusion

Need different models for different partnerships:
- Correct model will depend on goals and scope of project.

Flexibility is important:
- Need to be able to tweak model based on evolving needs, financial realities, etc.

Answer questions early, such as:
- Who is responsible for what?
- What is our exit strategy?
- How will IP issues be handled?

Large, broad partnerships (e.g, NCDS) can **lead to more targeted partnerships and vice versa.**

Be prepared for much work and (hopefully) big payoffs.

Questions?

- More examples?

NCDS Lessons Learned—Year 1

Value

Challenging to show value in very early stages, especially with broad membership. But this is essential to build momentum.

Managing Needs and Expectations

Members have different needs and expectations that must be balanced; cannot play favorites, yet want to show value to all.

Membership Challenges

How to accommodate multiple units within universities and businesses? How to accommodate small businesses and entrepreneurs? How to diversify? How to nurture "evangelicals" who will recruit new members?

Structural

Structural detail vs. need to be flexible as organization grows and matures; legal structure and details take time; RENCI must provide early leadership, but transition to a member-driven organization is critical.

Extra Slides

For Background

Other Partnership Models: Kaminario

- In collaboration with vendor, designed and conducted detailed performance testing of Kaminario all flash array with VMWare*

- Leveraged RENCI infrastructure for further testing within a research project:
  - Tested as an iSCSI device for rapid VM provisioning using ExoGENI

- Tested as relational database for genomics processing

- Ongoing relationship: more testing and evaluation planned, more academic papers and technical reports.

Kaminario 6TB all flash array

RENCI maintains ownership of $100K in hardware

*Work produced a joint White Paper:
http://www.hpcwire.com/off-the-wire/kaminario-renci-announce-report-storage-hpvcs/

renci RESEARCH \ ENGAGEMENT \ INNOVATION

REACH NC-Elsevier Model

Benefits:

- Custom support to manage updates, troubleshooting, etc.

- First in line/ first right of refusal to test upgrades & technical features.

- Ability to 'lead the pack' and work directly with the Elsevier technical team to customize and improve the tool

- Ability to pull REACH NC data to for development of custom visualization and analysis tools> (example: NAVEX tool)

Benefits:

- Exemplary use-case of a large, statewide implementation of their product (largest in the world)

- "Free advertising" and spokespeople for product.

- New scripts/ technologies/ data sources developed by RENCI partners to enhance the tool

RESEARCH \ ENGAGEMENT \ INNOVATION

REACH NC Organizational Structure
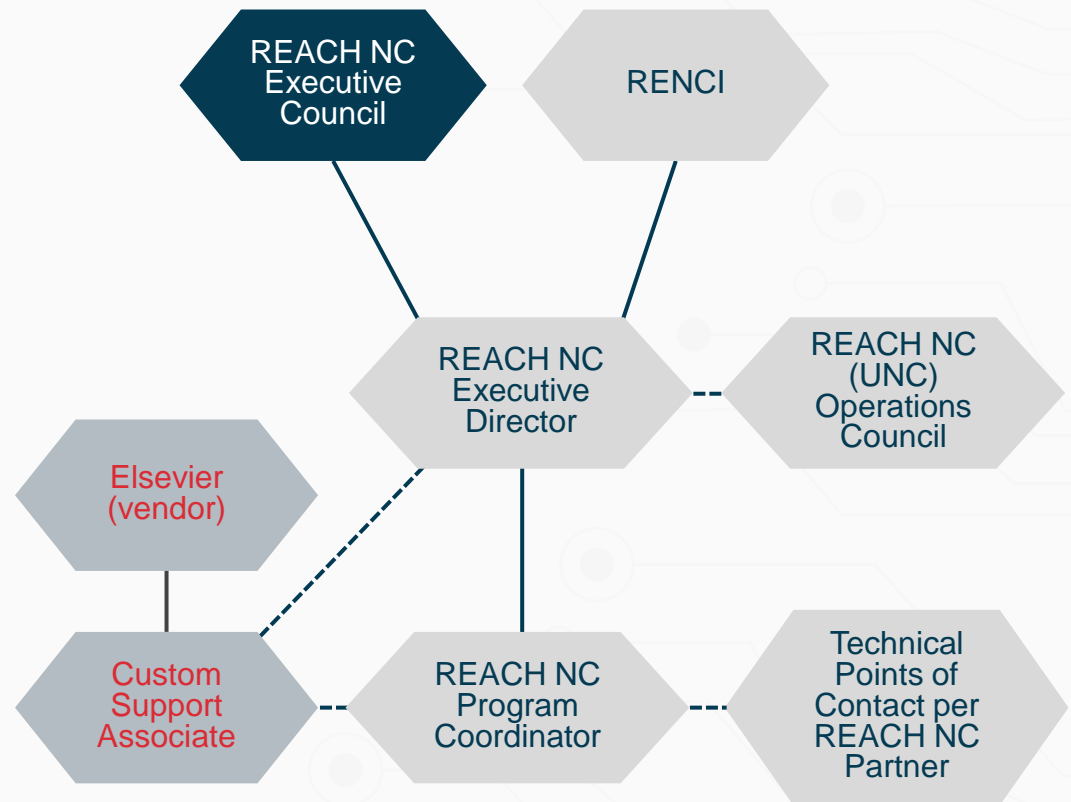
Internal (UNC) Operations Council

- Advise on technical and implementation-related issues.

Executive Council

- Advise on sustainability and long-term strategy for growth

Custom Support Associate

- Provided by vendor located at RENCI

RESEARCH \ ENGAGEMENT \ INNOVATION
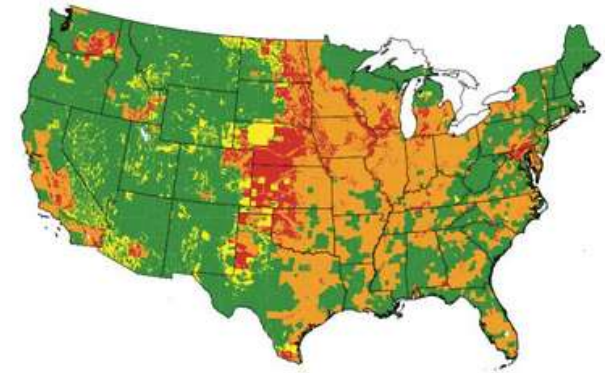
31

Start up! TerraHub

WHO NEEDS TERRAHUB?

If you rely on Google Maps or GIS for your business, TerraHub may be for you

Utility-scale Solar Power

Emergency Response

Policy, Govt & Business

IRODS Constortium Lessons Learned-Year 1

Flexibility is vital: **still figuring out if membership, support contracts, or service will be major revenue source**

Merging of flavors of iRODS into single version was correct

Middleware is difficult to describe:
Need Better Primers | Reference Installations | Interfaces as Sales Tools

- Need for iRODS technology in data centers is real and immediate.

- Hardware vendors may be a significant percentage of membership

renci
RESEARCH \ ENGAGEMENT \ INNOVATION