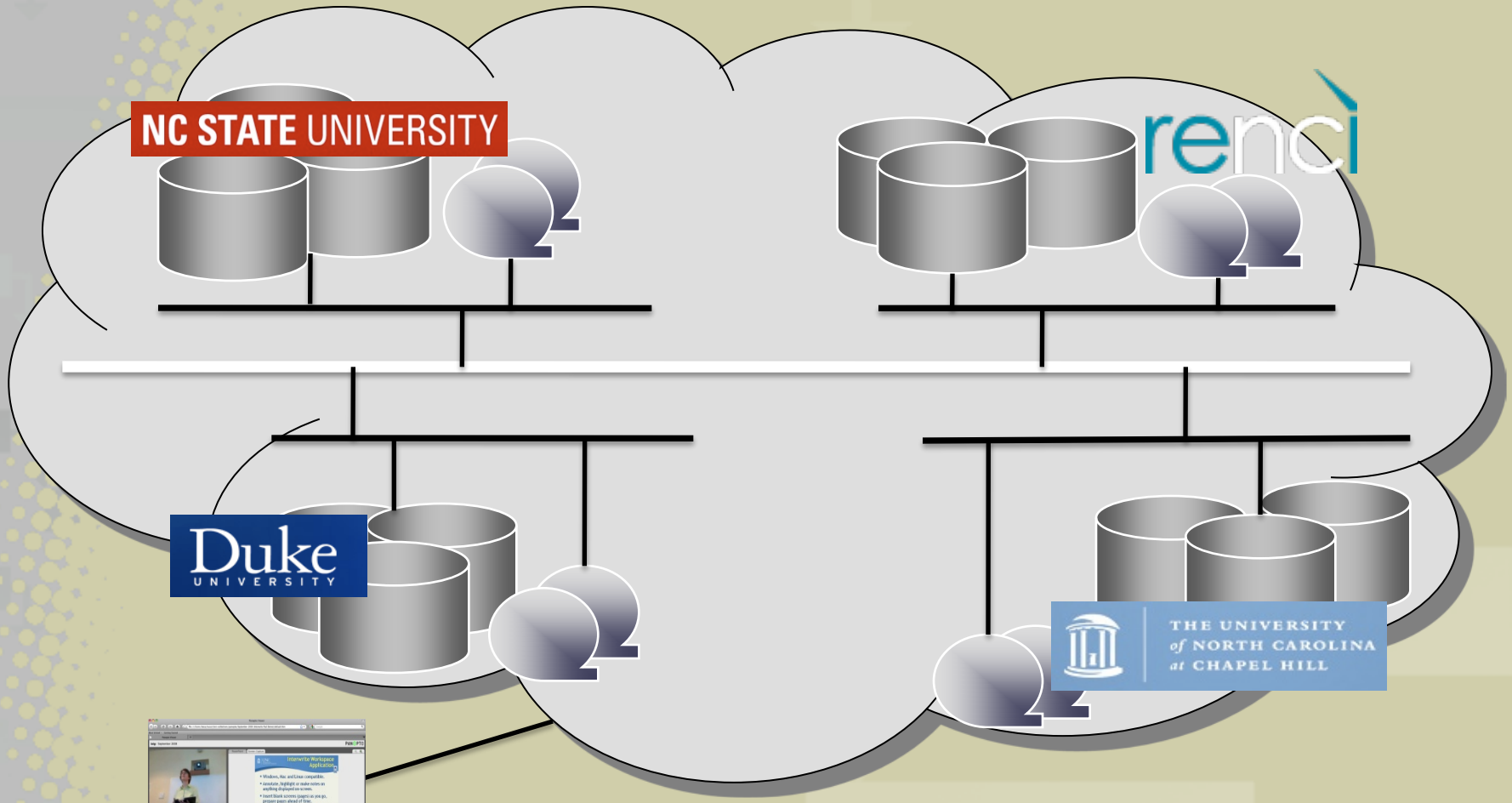
Managing Shared Digital Research Data in Federated Storage Clouds for Higher Education

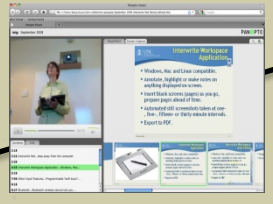
TUCASI data Infrastructure Project (TIP)

Richard J. Marciano

- A collaborative project of Duke, UNC, NC State, and RENCI
- Deployment of a prototype federated data infrastructure
- Leveraging data resources for competitive research and leadership
- A step toward a regional research data cloud

Federated Repositories





Funding Sources

- 2-year project: July 2009 – June 2011
- $2.7M pilot project
- Triangle Universities Center for Advanced Studies, Inc. (TUCASI), 1975
  - Established to ensure the continued presence of the research institutions in the Research Triangle Park
  - A 120-acre campus to house organizations that could bring together faculty from the three universities and Park scientists
- Project leverages earlier and ongoing funding by NSF/OCI, NARA and IMLS

Project Organization

- Project Lead: Richard Marciano (UNC/SALT)
- Project Manager: Amy Shoop (UNC ITS)
- Oversight Council
  - CIOs -- Head Librarians
    - Tracy Futhey -- Duke CIO — Deborah Jakubs -- Duke Librarian
    - Marc Hoit – NCSU CIO — Susan Nutter – NCSU Librarian
    - Larry Conrad – UNC CIO — Sara Michalak – UNC Librarian
  - RENCI
    - Alan Blatecky -- RENCI — Stan Ahalt -- RENCI
  - DICE Center
    - Reagan Moore – DICE
  - SALT Lab
    - Richard Marciano -- SALT

Focus Group Membership

Focus Groups	University Teams		
	Duke	**Chapel Hill**	**NC State**
Classroom Capture	**Samantha Earp (CC lead)** *(OIT-Academic Services)*	**Suzanne Cadwell** *(ITS-Academic Outreach & Engagement)* **Charlie Greene** *(ITS-Teaching & Learning)* **Pam Sessoms** *(Lib-e-Reference)*	**Lou Harrison** *(DELTA)* **Hal Meeks** *(OIT-Outreach, Communications and Consulting)*
Storage	**Amy Brooks** *(OIT-Systems)* **Klara Jelinkova** *(OIT-Shared Services & Infrastructure)* **David Kennedy** *(Lib-Info. Sys. Support)* **Molly Tamarkin** *(Lib-Systems)* **Jim Tuttle** *(Lib-Systems)*	**Reagan Moore (S lead)** *(DICE)* **Leesa Brieger** *(RENCI-Data)* **Brent Caison** *(ITS-Storage)* **Dave Pcolar** *(Lib-Systems)* **Bill Schulz** *(Lib-Systems)* **Lisa Stillwell** *(RENCI-Data)*	**Steve Morris** *(Lib-Systems)* **Eric Sills** *(OIT-Research Computing)*
Future Data & Policy	**Paolo Mangiafico** *(Provost-Dig. Info. Strategy)* **Tim Pyatt** *(Lib-Archives)*	**Ruth Marinshaw** *(ITS-Research Computing)* **Will Owen** *(Lib-Systems)* **Rich Szary** *(Lib-Special Collections)*	**Kristin Antelman (FD&P lead)** *(Lib)* **Susan Nutter** *(Lib-Head Librarian)*

TIP Goals and Accomplishments

- Provide common tools to allow seamless cross-site access
  - Fits with sites' heterogeneous infrastructure
  - Spans administrative diversity (local policies implemented)
  - Diverse data: research data, library resources, course capture
- Controlled data publication
  - Public data
  - Restricted data (varying levels of access permitted)
- Search and discovery portal: Search TRLN prototype
- Common authentication system (Shibboleth)
- Replication of data between sites
- Creation of policies for data deposit and access

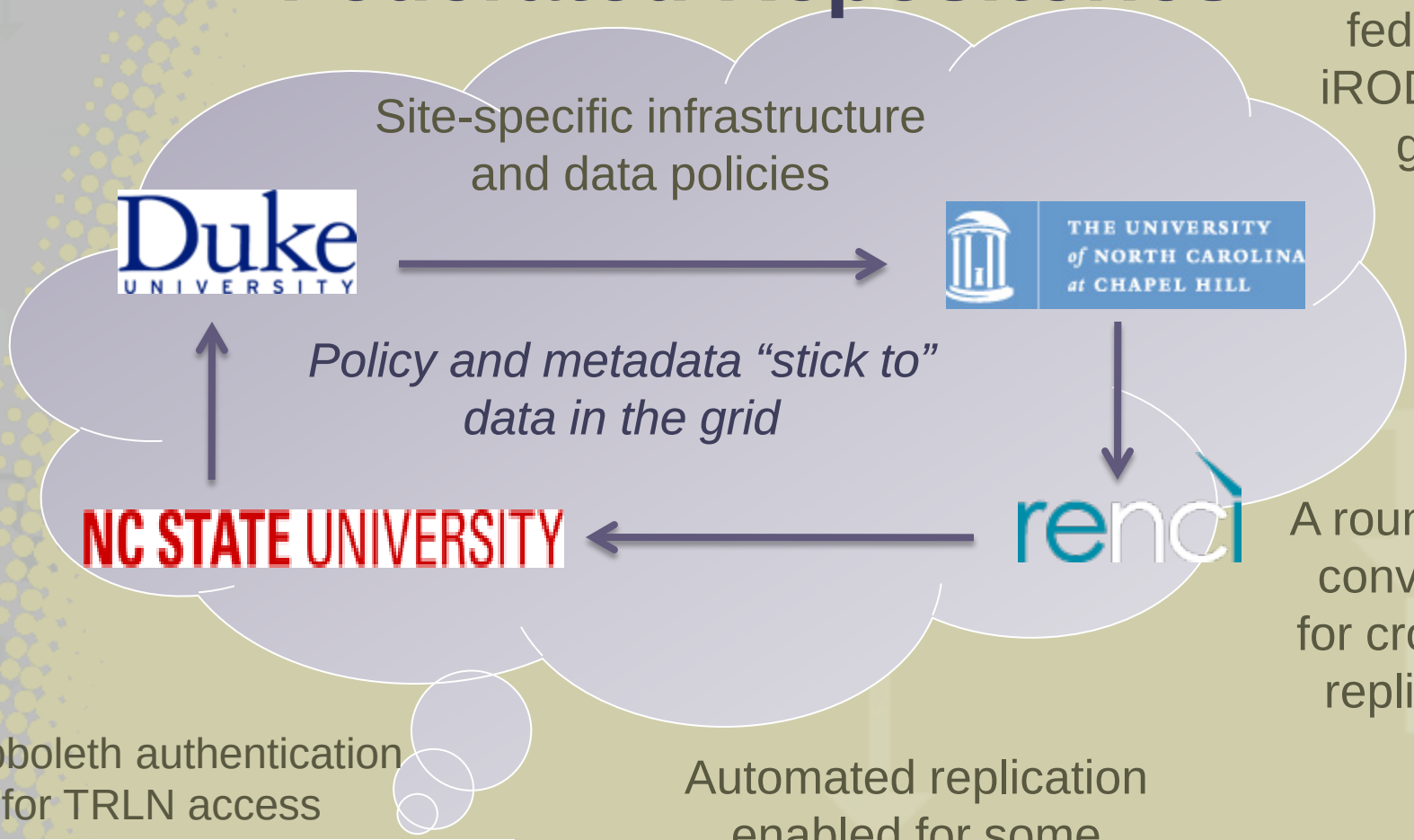
Cloud Services for Research

Data grids support interoperability across technologies

- manage name spaces for identifying records, archives, storage systems
- decouple access mechanisms from the storage system
- cross organizational, administrative and security boundaries
- details of retrieving data on each system handled by the grid

Discovery and Replication Across Federated Repositories

Four federated iRODS data grids



A round-robin convention for cross-site replication

Shibboleth authentication for TRLN access

Automated replication enabled for some collections

TIP components

- iRODS – Rule-Oriented Data System
  - Distributed Data Management
  - https://www.irods.org/pubs/iRODS_Fact_Sheet-0907c.pdf

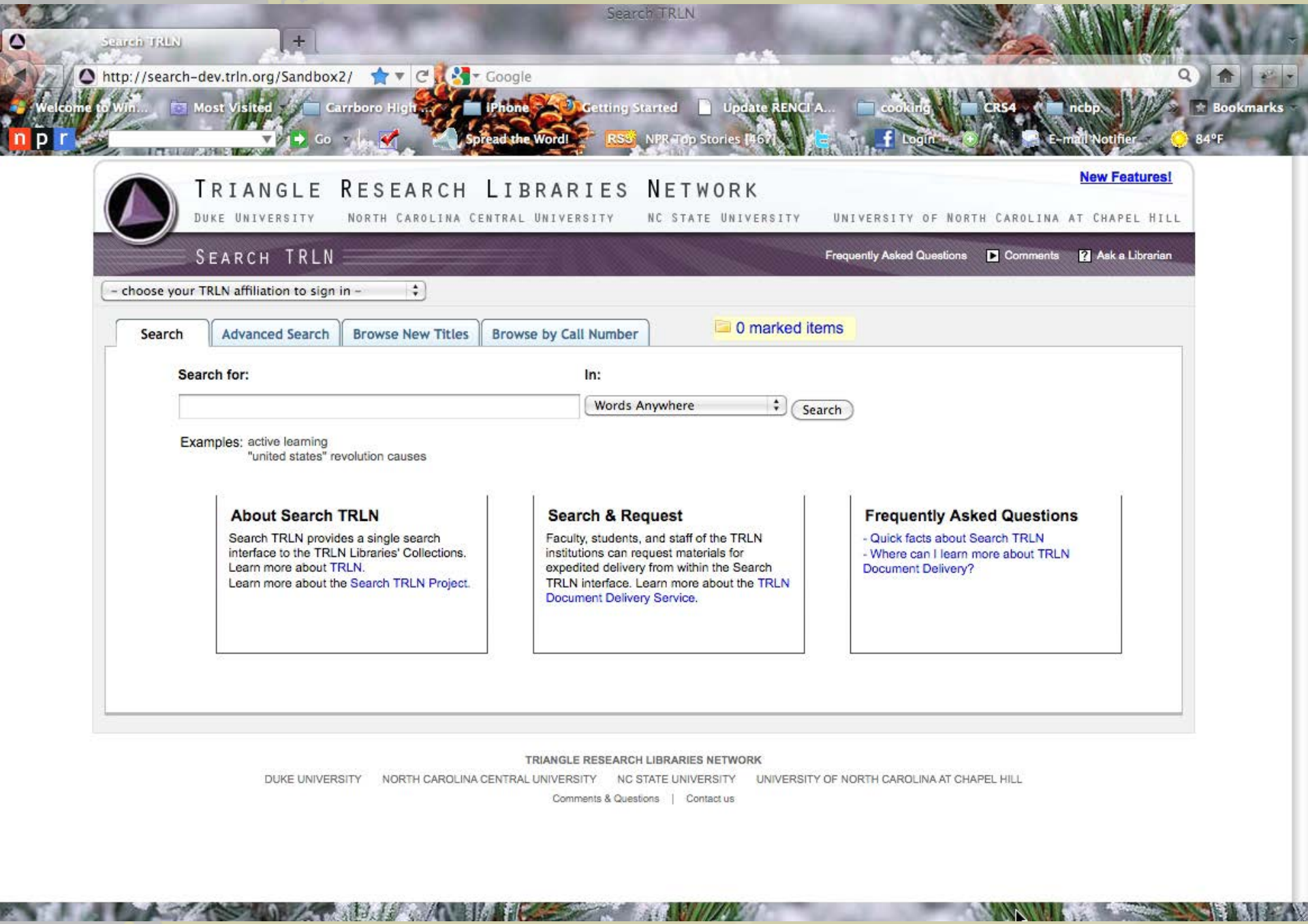
- Search TRLN
  - Federated Discovery Environment
  - http://search-dev.trln.org/Sandbox2/

TRIANGLE RESEARCH LIBRARIES NETWORK

- Shibboleth
  - Federated Single Sign-On
  - http://shibboleth.internet2.edu/about.html

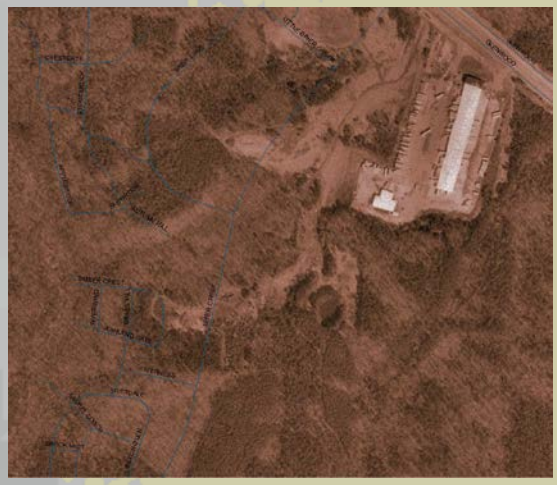
Access Methods for TIP Collections

- Web addressable content – SearchTRLN dev system
  - UNC North Carolina Collection - Digitized Postcard
  - Duke Classroom Capture
  - NCSU Color Digital Orthoimagery
- Web addressable content via iRODS
  - RENCI data access using Shibboleth


Search TRLN
Search TRLN
http://search-dev.trln.org/Sandbox2/
Google
Welcome to Win...
Most Visited
Carrboro High ...
iPhone
Getting Started
Update RENCI A...
cooking
CRS4
ncbp
Bookmarks
Go
Spread the Word!
NPR Top Stories [467]
Login
E-mail Notifier
84°F
TRIANGLE RESEARCH LIBRARIES NETWORK
New Features!
DUKE UNIVERSITY
NORTH CAROLINA CENTRAL UNIVERSITY
NC STATE UNIVERSITY
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
SEARCH TRLN
Frequently Asked Questions
Comments
Ask a Librarian
- choose your TRLN affiliation to sign in -
Search
Advanced Search
Browse New Titles
Browse by Call Number
0 marked items
Search for:
In:
Words Anywhere
Search
Examples: active learning
"united states" revolution causes
About Search TRLN
Search TRLN provides a single search interface to the TRLN Libraries' Collections.
Learn more about TRLN.
Learn more about the Search TRLN Project.
Search & Request
Faculty, students, and staff of the TRLN institutions can request materials for expedited delivery from within the Search TRLN interface. Learn more about the TRLN Document Delivery Service.
Frequently Asked Questions
- Quick facts about Search TRLN
- Where can I learn more about TRLN Document Delivery?
TRIANGLE RESEARCH LIBRARIES NETWORK
DUKE UNIVERSITY
NORTH CAROLINA CENTRAL UNIVERSITY
NC STATE UNIVERSITY
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
Comments & Questions
Contact us

NCSU - Brier Creek time series imagery

1993



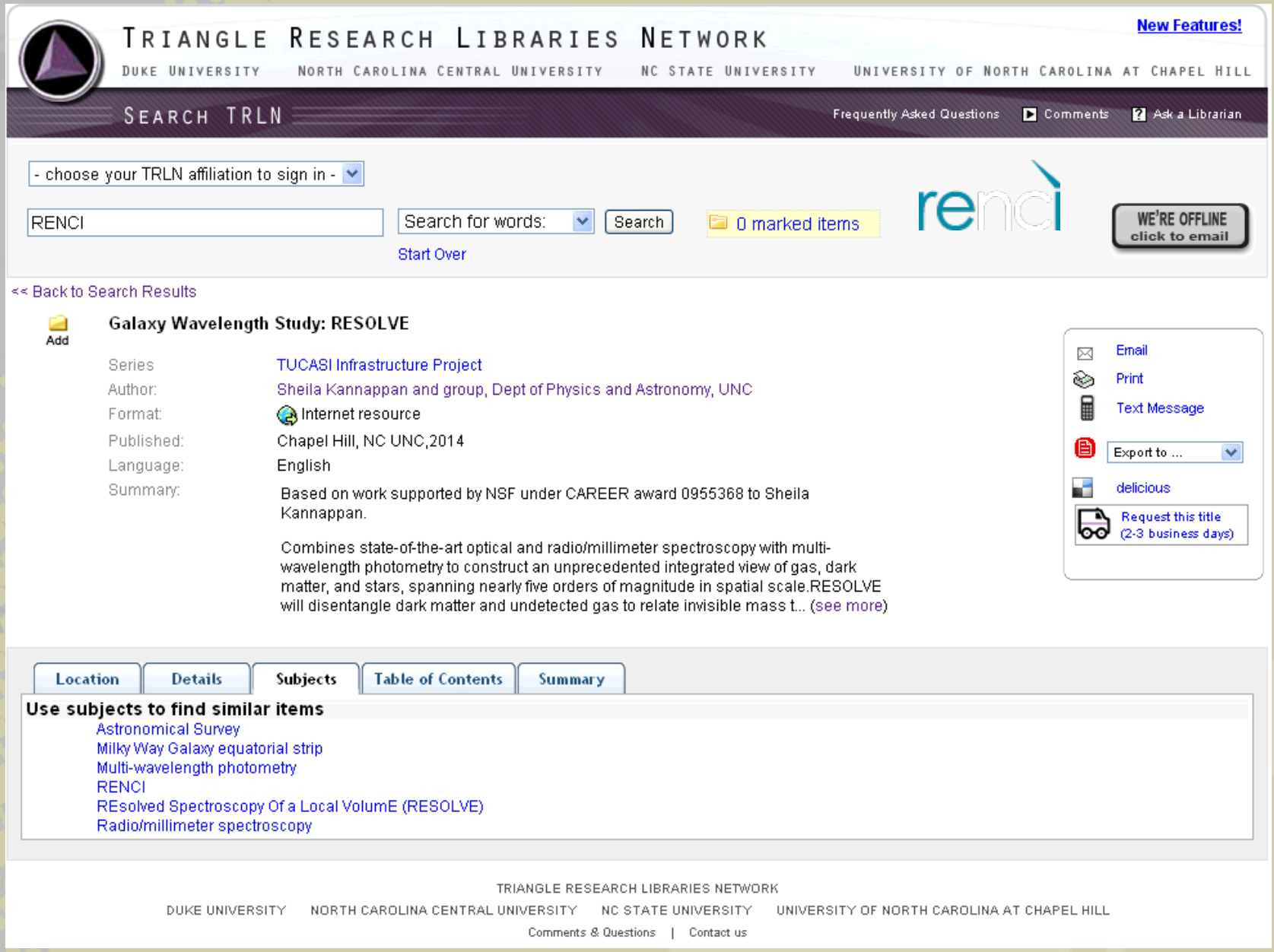
1998

1999

2002

2005

Use case:
Land use and impervious surface change analysis


TRIANGLE RESEARCH LIBRARIES NETWORK
New Features!
DUKE UNIVERSITY NORTH CAROLINA CENTRAL UNIVERSITY NC STATE UNIVERSITY UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
SEARCH TRLN
Frequently Asked Questions
Comments
Ask a Librarian
- choose your TRLN affiliation to sign in -
RENCI
Search for words:
Search
0 marked items
Start Over
renci
WE'RE OFFLINE
click to email
<< Back to Search Results
Add
Galaxy Wavelength Study: RESOLVE
Series TUCASI Infrastructure Project
Author: Sheila Kannappan and group, Dept of Physics and Astronomy, UNC
Format: Internet resource
Published: Chapel Hill, NC UNC,2014
Language: English
Summary: Based on work supported by NSF under CAREER award 0955368 to Sheila Kannappan.
Combines state-of-the-art optical and radio/millimeter spectroscopy with multi-wavelength photometry to construct an unprecedented integrated view of gas, dark matter, and stars, spanning nearly five orders of magnitude in spatial scale.RESOLVE will disentangle dark matter and undetected gas to relate invisible mass t... (see more)
Email
Print
Text Message
Export to ...
delicious
Request this title
(2-3 business days)
Location
Details
Subjects
Table of Contents
Summary
Use subjects to find similar items
Astronomical Survey
Milky Way Galaxy equatorial strip
Multi-wavelength photometry
RENCI
REsolved Spectroscopy Of a Local VolumE (RESOLVE)
Radio/millimeter spectroscopy
TRIANGLE RESEARCH LIBRARIES NETWORK
DUKE UNIVERSITY NORTH CAROLINA CENTRAL UNIVERSITY NC STATE UNIVERSITY UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
Comments & Questions | Contact us

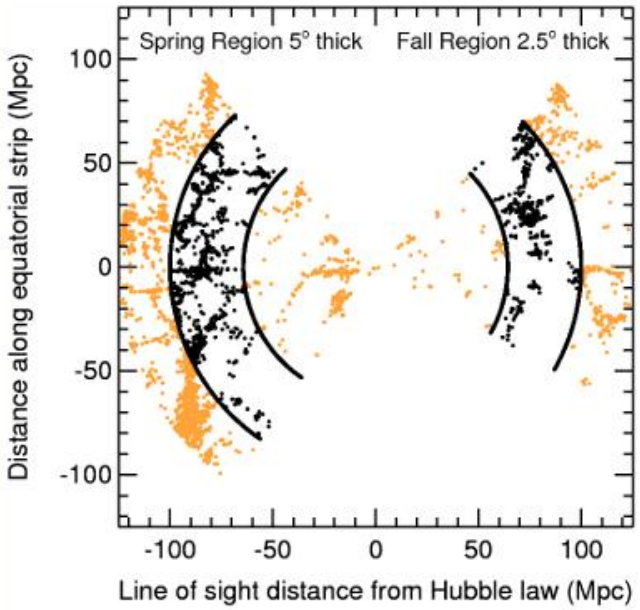
RESOLVE

REsolved Spectroscopy Of a Local VolumE

RESOLVE is a volume-limited census of stellar, gas, and dynamical mass as well as star formation and merger activity for all galaxies and larger structures, from dwarf galaxies of baryonic mass ~10^9 $M_\odot$ up to groups, filaments, and voids on tens of Mpc scales, in 53,000 cubic Mpc of the nearby universe. The survey's science drivers include relating galaxy velocity/mass functions to environment, constraining the "missing baryons" problem from a complete accounting perspective, and understanding disk building in large-scale context. See our outreach page for a non-technical description.

Survey Area

The RESOLVE Survey (black points) shown within a portion of the larger SDSS Redshift Survey from which it was drawn. RESOLVE comprises ~1500 galaxies in two equatorial strips (RA 22-3 hr, Dec -1.25 to +1.25 in the fall; RA 8.75-15.75, Dec 0 to +5 in the spring) at redshifts 4500-7000 km/s. The fall strip largely coincides with SDSS Stripe 82.



Movie Time…

- A quick fly-through of the interface:
  - 3 min 39 sec

Implementation Issues

- Establishment of Data Policy is crucial
  - cross-site, inter-institutional
  - data access and modification policies
  - preservation and curation (data life cycle evolution)
- Researcher-technologists and librarian-archivists *together* provide best use/curation policies and implementations
- Adequate personnel support is essential to turning hardware into useful, performant infrastructure

TIP infrastructure: a model approach?
NSF/NIH/NEH Data Management

- Requires researchers to define data policy
- Requires support from professionals in data management (librarians): *preservation* principles, standards, engineering, technology, and management
- Requires institutional support:
  - storage space
  - support for sharing and publishing data
  - infrastructure for policy support: cross-site collaborations, site-specific administration policies, storage systems, naming conventions, etc.

Future Uses of the Infrastructure

Widening the Context of the Data Use

- Research Data
  - Astronomy: publishing data and educational services
  - Genomics: private data and locally-stored public data
  - NC geospatial data: local copies and derived data products
  - Social Sciences: data analysis and visualization tools
- Libraries:
  - Preservation and Access: Carolina Digital Repository
  - GIS Discovery and Geospatial Service Framework
- Instruction:
  - Course Capture
  - Online Learning