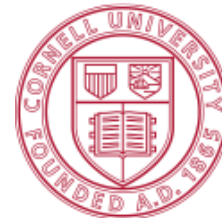# The National Data Service: A Library Perspective

Dean B. Krafft

Chief Technology Strategist, Cornell University Library

NDS Consortium Planning Workshop

June 12-13, 2014

Cornell University Library

Anne R. Kenney,
Carl A. Kroch
University
Librarian, Cornell
University

Platitudes for the NDS

- Define your niche (don't be everything for everyone)
- Play well with others (don't be a silo)
- Leverage existing systems and solutions (ORCID, DataCite, SHARE, VIVO, DPN)
- Meet the needs of researchers, research institutions, research centers, CASC members, and funders
- Create a believable business model
- Use Linked Open Data to represent how datasets relate to the entire research ecosystem

Libraries' Role in the National Data Service

What can libraries do?

- Help researchers describe their data
- Help them pick formats to store it in
- Provide guidance and services around choosing repositories and providing access
- Provide guidance on privacy, licensing, and sharing issues
- Ensure preservation of appropriate data for future research reuse

What more can libraries do?

- Collaborate at scale across institutions and disciplines
- Help link the data with its research context to make it more discoverable and reusable
- Help link it to publications about the research
- Provide collaborative tools and spaces for researchers to work with the data
- Provide the people and organizational support to help researchers to manage research data

What Can't Libraries Do?

- Handle the Data Deluge – "really big data"
- Fund this ourselves – we need a business model to support the costs
- Do work that doesn't clearly benefit our own researchers and institutions
- Provide cyberinfrastructure to support analysis, simulation, and visualization
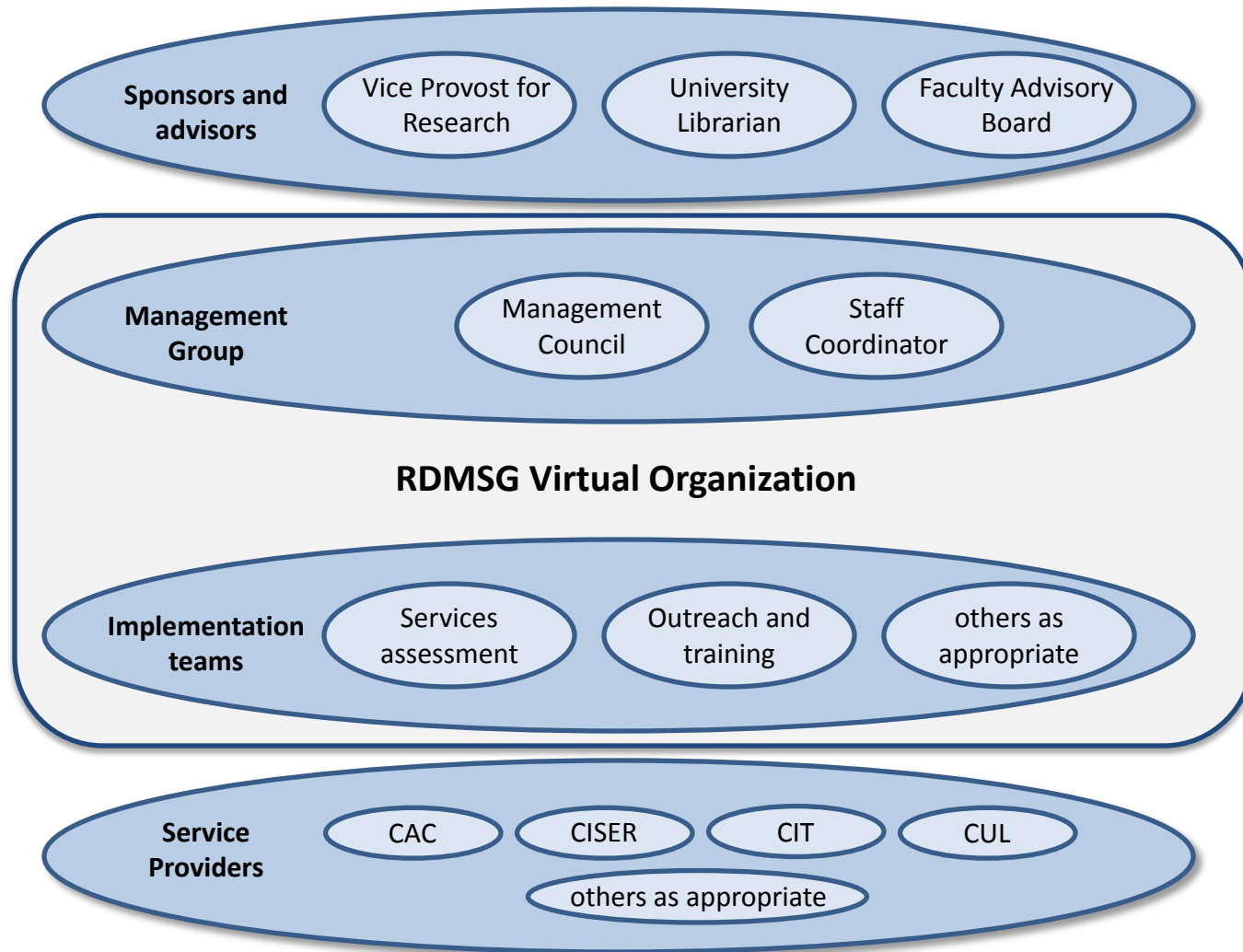
Research data at scale raises different issues than libraries have dealt with in the past. Libraries must **evolve** to meet the needs of tomorrow's scholars and researchers.

J.W. Audubon - Ivory-billed Woodpecker

What does the Cornell Library provide now?

- Cornell's Institutional Repository (IR) – originally designed for journal articles
- Already used for research data deposit
- Added new policy and support:
  - Up to 10Gigabytes/year per research project
  - Recommendations on metadata
  - Recommendations on licensing – generally suggest Open Data Commons licenses (preferably the Public Domain Dedication and License – PDDL): http://opendatacommons.org/licenses/

Research Data Management Service Group (RDMSG)

Sponsors and advisors
- Vice Provost for Research
- University Librarian
- Faculty Advisory Board

RDMSG Virtual Organization

Management Group
- Management Council
- Staff Coordinator

Implementation teams
- Services assessment
- Outreach and training
- others as appropriate

Service Providers
- CAC
- CISER
- CIT
- CUL
- others as appropriate

Cornell RDMSG Services

- Consultation on Data Management Planning
- Maintain "Best Practices" for Data Management
- Provide specific advice and pointers to services for:
  - Collaboration
  - Data analysis
  - Data sharing
  - High performance computing
  - Intellectual property and copyright
  - Metadata
  - Privacy and confidentiality
  - Storage, backup, and recovery

http://data.research.cornell.edu

Cornell University Library

Search Library   Search Cornell

arXiv.org

Search or Article-id   (Help | Advanced search)

All papers   Go!

Open access to 549,302 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics

Subject search and browse: Physics   (Search) (Form Interface) (Catchup)

8 Apr 2009: Added public author identifiers, Facebook interaction, myarticles widget, and personal Atom feeds
31 Mar 2009: Quantum Gases (cond-mat.quant-gas) subject area added
See cumulative "What's New" pages.
Robots Beware: indiscriminate automated downloads from this site are *not* permitted.

Physics

- Astrophysics (**astro-ph** new, recent, find)
  includes: Cosmology and Extragalactic Astrophysics; Earth and Planetary Astrophysics; Galaxy Astrophysics; High Energy Astrophysical Phenomena; Instrumentation and Methods for Astrophysics; Solar and Stellar Astrophysics
- Condensed Matter (**cond-mat** new, recent, find)
  includes: Disordered Systems and Neural Networks; Materials Science; Mesoscale and Nanoscale Physics; Other Condensed Matter; Quantum Gases; Soft Condensed Matter; Statistical Mechanics; Strongly Correlated Electrons; Superconductivity
- General Relativity and Quantum Cosmology (**gr-qc** new, recent, find)
- High Energy Physics - Experiment (**hep-ex** new, recent, find)
- High Energy Physics - Lattice (**hep-lat** new, recent, find)
- High Energy Physics - Phenomenology (**hep-ph** new, recent,
- High Energy Physics - Theory (**hep-th** new, recent, find)
- Mathematical Physics (**math-ph** new, recent, find)
- Nuclear Experiment (**nucl-ex** new, recent, find)
- Nuclear Theory (**nucl-th** new, recent, find)
- Physics (**physics** new, recent, find)
  includes: Accelerator Physics; Atmospheric and Oceanic Phys
  Chemical Physics; Classical Physics; Computational Physics; [
  Geophysics; History of Physics; Instrumentation and Detector
  Physics; Popular Physics; Space Physics
- Quantum Physics (**quant-ph** new, recent, find)

E-Prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics

- \>944,498 e-prints in the repository
- 5-6 million articles downloaded each month
- 100,000 distinct users per day worldwide
- Registered subscribers live in 200 countries

arXiv.org

Data associated with this Submission

Start >> **Add Files** >> **Process** >> **Metadata** > **Add Data** >> **Preview**

File uploaded

Here you may upload data sets, imagery, simulations, videos, or other material of substantial file size (up to 100MB per file, 1GB total), which is of relevance to your submission while exceeding the size restrictions imposed by arXiv. These files will be relayed to the Data Conservancy Project (DC) for archival storage and access.

There will be mutual links from arXiv to DC to connect the external material with this submission. Visit arXiv's dataset help for more information.

Data uploaded here will be distributed under the same license as the article. You may change the license on the [Start] screen.

! Material uploaded in this section will be relayed to the Data Conservancy. It will remain associated with your submission, but it will not be stored at arXiv.

This is a pilot project which explores future cooperation between arXiv and DC and the related infrastructure requirements. As such this service may undergo substantial changes and is not guaranteed to persist beyond the duration of the pilot. However reasonable efforts will be made to keep the uploaded material accessible.

Add data

Choose File | No file chosen

Upload file

*Short description

Continue: Preview

Data Files Attached to this Submission			Delete All
Filename	**Type**	**Description / SHA256 checksum**	**Actions**
spectra_x34g56.dat	unknown	Spectra readout for sample x34g56 2629a6a11e9a0f908a71222f7b6d746aa2b2c3e2078cc9716b4e3bec9c6880ab	Delete

arXiv.org > cond-mat > arXiv:1104.5494

Search or Article-id

(Help | Advanced search)

All papers ▼ Go!

Condensed Matter > Strongly Correlated Electrons

Thermodynamics of Strongly Interacting Fermions in 2D Optical Lattices

Ehsan Khatami, Marcos Rigol

(Submitted on 28 Apr 2011)

We study finite temperature properties of strongly correlated fermions in two-dimensional optical lattices by means of numerical linked cluster expansions. Our results are exact in the thermodynamic limit and we focus our analysis on the strongly interacting regime, where the on-site repulsion is of the order or greater than the bandwidth. We compute the equation of state, double occupancy, entropy, and spin correlations for temperatures that are similar or below the ones achieved in current optical lattice experiments. A cooling scheme for those experiments is also discussed.

Comments: 5 pages, 6 figures
Subjects: **Strongly Correlated Electrons (cond-mat.str-el)**
Cite as: **arXiv:1104.5494v1** [cond-mat.str-el]

Submission history

From: Ehsan Khatami [view email]
[v1] Thu, 28 Apr 2011 20:00:02 GMT (583kb,B)

Which authors of this paper are endorsers?

Link back to: arXiv, form interface, contact.

Download:

- PDF
- PostScript
- Other formats

Data sets (what is this?)

- Data Conservancy (217 files)

Current browse context:

cond-mat.str-el
< prev | **next >**
new | recent | 1104

Change to browse by:

cond-mat

References & Citations
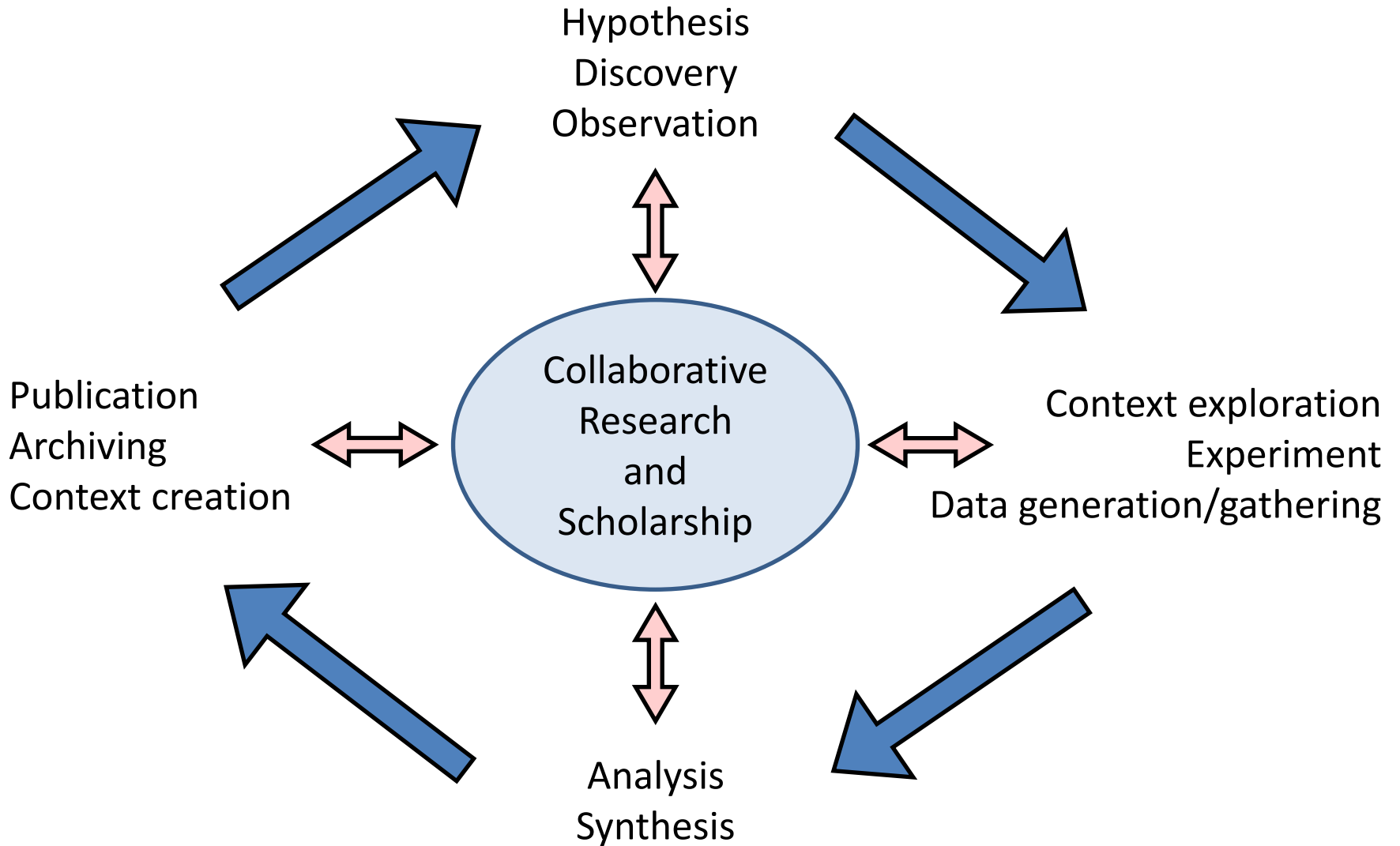
- NASA ADS

Bookmark (what is this?)

The National Data Service's Role in the Data Ecosystem

The Cycle of Research

Hypothesis
Discovery
Observation

Publication
Archiving
Context creation

Collaborative
Research
and
Scholarship

Context exploration
Experiment
Data generation/gathering

Analysis
Synthesis

SHared Access Research Ecosystem (SHARE)

- Collaboration of ARL, AAU, and APLU
- "a higher education and research community initiative to ensure the preservation of, access to, and reuse of research outputs"
- Response to OSTP open access mandate
- Applies to all research outputs (including data)
- Step 1: Notification system (think Twitter)
- Step 2+: A distributed content and registry layer that supports discovery and research data

Digital Preservation Network (DPN)
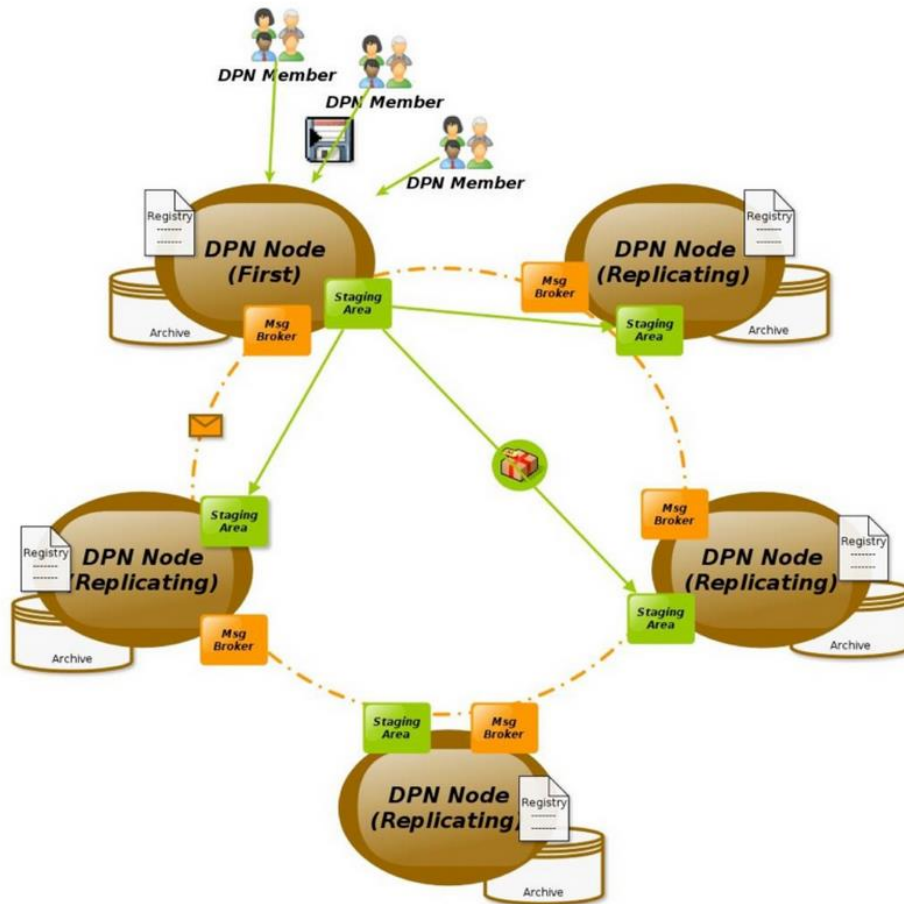
1. Establishes a network of heterogeneous, interoperable, trustworthy, preservation repositories

2. Replicates content across the network, to multiple nodes

3. Enables restoration of preserved content to any node in the event of data loss, corruption or disaster

4. Ensures the ongoing preservation of digital information from depositors in the event of dissolution or divestment of depositors or repository(ies)

5. Provides the option to (technically and legally) "brighten content" preserved in the network over time

http://www.dpn.org

Other Systems and Solutions

- ORCID (orcid.org): a LOD-compatible international standard for identifying researchers

- DataCite (datacite.org): Using Digital Object Identifiers (DOIs) to promote citation of research data

- Research Objects (researchobject.org): Focused on creating an aggregation object that bundles together experimental resources that are essential to a computational scientific study or investigation (uses OAI-ORE annotations)

Linked Open Data, VIVO, and the Structured Representation of Research and Scholarship

VIVO connects scientists and scholars with and through their research and scholarship
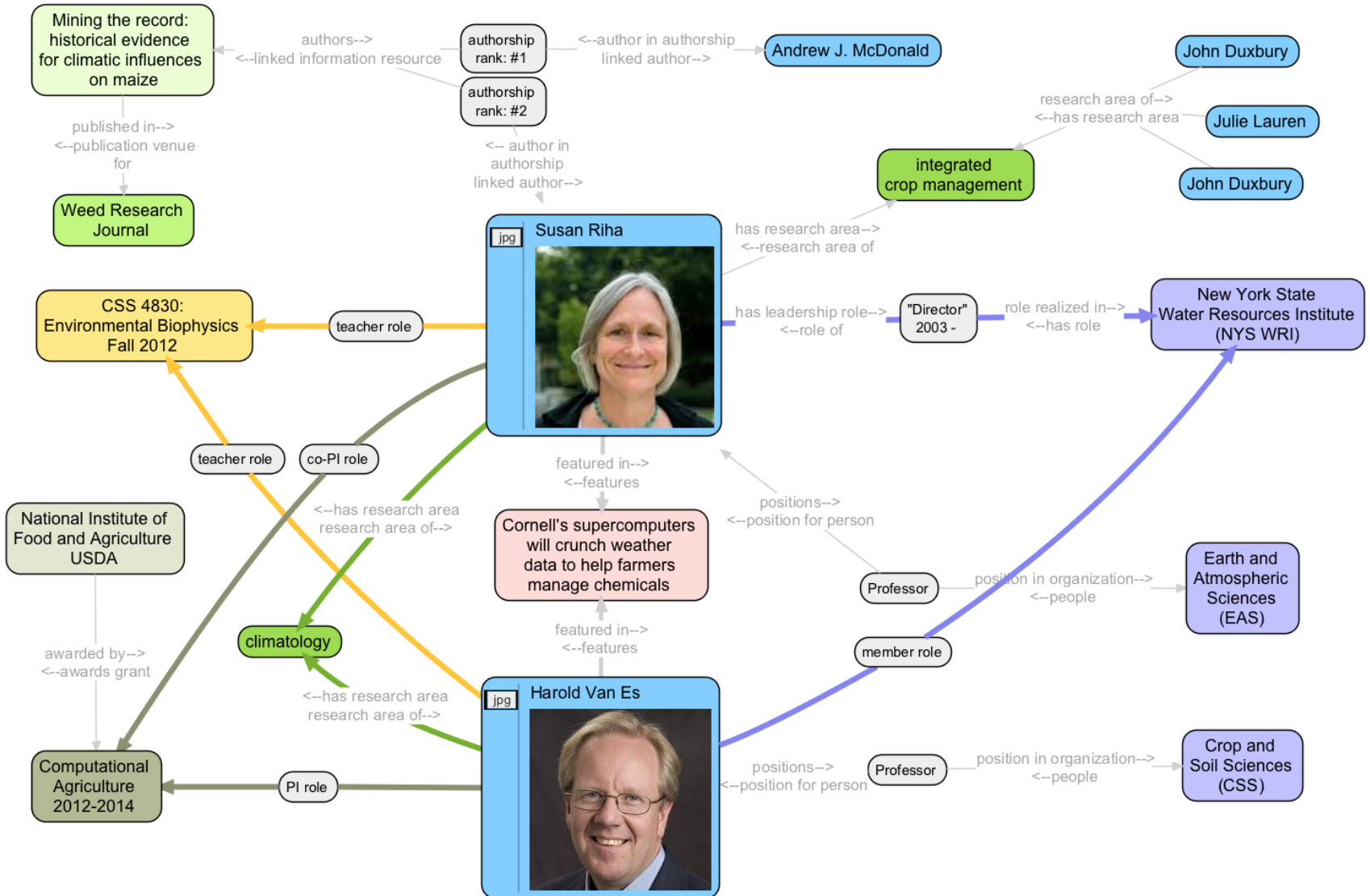
What is VIVO?

- Software: An open-source semantic-web-based researcher and research discovery tool
- Data: Institution-wide, publicly-visible information about research and researchers
- Standards: A standard ontology (VIVO data) that interconnects researchers, communities, and campuses using Linked Open Data
- Community: An open community with strong national and international participation, with a legal and financial home in DuraSpace, led by the VIVO Project Director: Layne Johnson

Linked data indexing for search

ICTS | INSTITUTE *for* CLINICAL & TRANSLATIONAL SCIENCE
AT THE UNIVERSITY OF IOWA
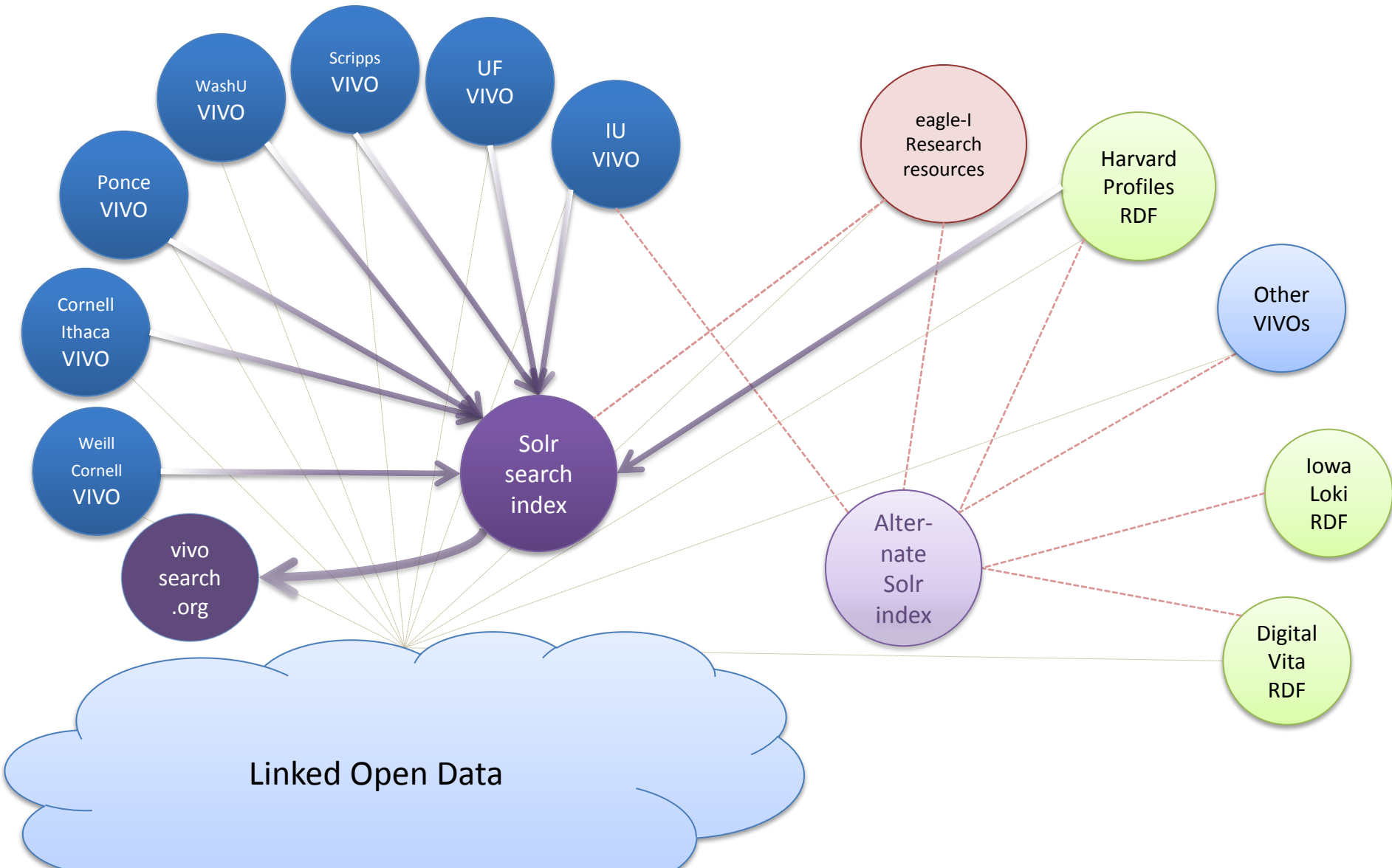
CTSAsearch Home

CTSA Search

CTSA Community Map

CTSA Institution Map

Participation Details

Custom Maps

Chicago Women in STEM Map

Comparison Approaches

Google Search

Federated Search

What is CTSAsearch?

CTSAsearch is a prototype demonstrating federated search using Linked Open Data published by members of the CTSA Consortium and other interested parties. To try it out, use the form below or click on the "CTSA Search" entry in the menu on the left to see a ranked list of matching investigators. Use the second form or click on the "CTSA Map" entry in the menu to visualize coauthorship amongst the matching investigators.

Search for Investigators at Multiple Institutions

☐ Display map

○ Text only

○ Text and UMLS concepts

● UMLS concepts (including support for boolean search using &,|, and !)

[   ] [Search]

Current Status

⁙ SPARQL endpoints queried: 10

⁙ Institutions indexed: 19

⁙ Total persons indexed: 124,945

⁙ Total publications by those persons indexed as part of their profile: 1,325,716

⁙ Total co-author pairs (two people on the same paper): 2,036,131

http://research.icts.uiowa.edu/polyglot/

Adding Research Resources and Facilities to VIVO

- CTSAconnect
  - OHSU, Harvard, Cornell, Florida, Buffalo & Stony Brook
  - eagle-i sister NIH project – Harvard, OHSU, 7 others
- Facilities, services, techniques, protocols, skills, and research outputs beyond publications
  - Extended ways to represent expertise
  - Improve attribution for data and other contributions to science

eagle-i inventories "invisible" resources

Research generates many resources that are rarely shared or published:

Biological Specimen (2935)

Human Study (250)

Instrument (3961)

Organism or Virus (28698)
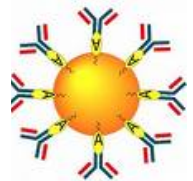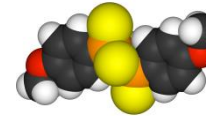
Protocol (630)

Reagent (6832)

Research Opportunity (54)

Service (2592)

Software (802)

Core Laboratories (466)

Over 47,000 resources currently in eagle-i

*All of these things are provided by or generated by a person or an organization

Connecting researchers, resources, and clinical activities

Deep Carbon Observatory Data Science Platform

LASP
Laboratory for Atmospheric and Space Physics

About   **Missions & Projects**   **Science**   **Engineering**   **Mission Operations & Data Systems**   **Education**

### Mission Operations & Data Systems

- Flight Operations
- Operational Software
- Planning & Scheduling
- Data Systems
- Data Products
- Facilities
- Personnel
- Mission Highlights

Data Products

LASP generates, maintains, and disseminates a range of space and atmospheric science data, including both individual data sets and broad data collections. Available data sets span many scientific disciplines and more than four decades of space-based research at LASP, and are summarized below. For data user assistance, please see the contact information within each data resource below.

- Key LASP Data Sets
- Data Collections
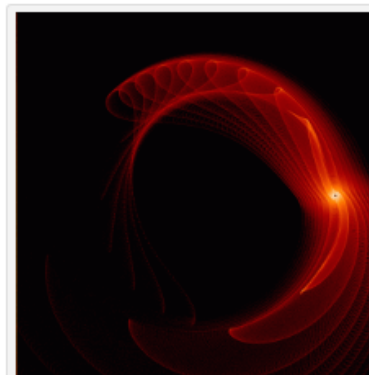
A computer model generated this image of atmosphere escaping Jupiter. (Courtesy LASP)

Related

- Mission Ops & Data Features & News
- Missions and Projects
- CU Students at LASP
- Staff Directory

Key LASP Data Sets

Name	Parameter(s)	Observable	Temporal coverage
Solar Dynamics Observatory (SDO) EUV Variability Experiment (EVE)	Soft X-ray and Ultraviolet Spectral Solar Irradiance (SSI)	Sun	2010-present
Solar Radiation and Climate Experiment (SORCE)	Total Solar Irradiance (TSI); Soft X-ray, UV, Visible, and near-Infrared Solar Spectral Irradiance (SSI)	Sun	2003-present
Aeronomy of Ice in the Mesosphere (AIM) Cloud Imaging and Particle Size instrument (CIPS)	Ultraviolet albedo; Polar Mesospheric cloud images, ice water content, particle radius	Earth	2007-present
Thermosphere Ionosphere	Soft X-ray and Ultraviolet solar	Sun	2002-present

Look to feed these dynamically rather than statically

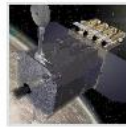# LASP

Flight Equipment

Instrument (42)

▸ Space Craft (8)

Space Craft

▸All A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
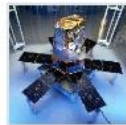

[Cassini Orbiter](#)

[POLAR](#)

[Solar Dynamics Observatory (SDO)](#)

[Solar Mesosphere Explorer (SME)](#)

[Solar Radiation and Climate Experiment (SORCE)](#)

[Student Nitric Oxide Explorer (SNOE)](#)

CU | LASP

Search

Home | Flight Equipment | People | Research | Storage

Woods, Thomas N | Associate Director of T... Divisions

Dr. Tom Woods is Associate Director of Technical Divi... Laboratory for Atmospheric and Space Physics (LASP) University of Colorado. He obtained his BS in Physics i... from Southwestern at Memphis (now Rhodes College) in Physics in 1985 from the Johns Hopkins University direction of Dr. Paul Feldman. Tom joined LASP in 198... on the UARS SOLSTICE program under the direction of Rottman. He originally served as the SORCE Project Sc... more)

Research Areas

Astrophysics | Solar Physics

Affiliation | Publications

Affiliation

Principal Investigator Of

Extreme Ultraviolet Variability Experiment (EVE)

Solar Extreme Ultraviolet Experiment (SEE)

Solar Radiation and Climate Experiment (SORCE)

Solar Stellar Irradiance Comparison Experiment (SOLSTICE)

CU | LASP

Search

Home | Flight Equipment | People | Research | Storage

Total Irradiance Monitor (TIM) | Instrument 🔗

Overview | Publications | Other

Overview

Description

The Total Irradiance Monitor (TIM) measures the total amount of radiation coming from the Sun. The sensor uses what is known as an absolute radiometer and houses four cone-shaped cavities. One of the cavities has an oscillating shutter that allows direct sunlight to shine into one of the cones. The material in the cone absorbs nearly all the Sun's energy and heats up. By measuring the voltage needed to bring this heated cone back to the same temperature as one of the other "reference" cones, which are kept at a constant temperature, the instrument can obtain an extremely accurate reading of the TSI in watts.

Is an Instrument on

SORCE (January 25, 2003 – Present)

Publications

supported publications

A new, lower value of total solar irradiance: Evidence and climate significance

Intercomparison of SCIAMACHY and SIM vis-IR irradiance over several solar rotational timescales

Solar total irradiance in cycle 23

Closing Thoughts

- What will compel researchers to share data and use the NDS?

- How does the NDS move from the vision of a few to a broad community effort?

- What unique services does the NDS provide that are so compelling that we HAVE to be part of it?

- If the NDS succeeds, does (your organization here) win too?

Questions?