

Cornell University

Social Science Gateway
to Teragrid
@ Cornell University

(also known as VirtualRDC)

Social Science Gateway provides:

- Usual social science data management and analysis tools not available on TeraGrid (TG) itself
 - SAS, SPSS, Stata
- Also provides standard TG tools
 - R, compilers, etc.
- Graphical desktop access to SSG compute nodes

Social Science Gateway provides (cont)

- High-speed (10Gbps) access to TeraGrid and thus to thousands of compute nodes and/or data
- Local data storage (base allocation includes some storage, additional storage from Cornell storage cloud at cost)

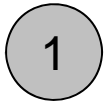
Typical user

- Researcher has existing theory/model/data
- Project requires substantially more resources than typically locally available
- Researcher
 - is comfortable with the traditional data tools (SAS, SPSS)
 - Less comfortable with TeraGrid tools

Typical workflow

- (1) Project inception
- (2) Researcher requests TeraGrid allocation for full computational resources, obtains TG-wide credentials
- (3) Researcher requests Social Science Gateway (SSG) access with existing TG-wide credentials

Project Workflow



2 Researcher requests access to TeraGrid

The screenshot shows the TeraGrid website homepage. At the top, there is a navigation bar with the TeraGrid logo on the left and a search bar on the right. Below the logo are navigation buttons for 'About', 'News', 'Outreach', 'Science Gateways', and 'User Support'. The search bar contains the text 'Search TeraGrid' and a search icon. Below the navigation bar, there are two main content sections: 'News & Highlights' and 'For Our Users'. The 'News & Highlights' section features a large banner for 'TERAGRID '09' held in Arlington, Virginia, from June 22-25, 2009. Below the banner, there is text describing the conference and providing information on registration and funding. The 'For Our Users' section includes a 'User Portal' login form with fields for 'Username:' and 'Password:', and a 'Login' button. Below the login form, there is a list of links for users, including 'XD Transition Information', 'Data and Computation Resources', 'Getting Started Guide', 'User & System News', 'Campus Champions Support Local Users', 'Search the Knowledge Base', 'Get a TeraGrid Allocation', 'Email Helpdesk/Support', 'Learn About Science Gateways', and 'Explore Advanced Support for TeraGrid Applications'. At the bottom of the page, there are two more sections: 'Science Highlights' and 'Calendar of Events'.

News & Highlights

TERAGRID '09

June 22-25, 2009 Arlington, Virginia

The 4th annual TeraGrid Conference (TeraGrid '09) will showcase presentations, posters, visualizations, and more. Attendees will learn how to leverage TeraGrid resources and services to achieve maximum results for their research, education, and innovation.

Reserve your room now at the [Hyatt Regency Crystal City](#). Special TeraGrid '09 conference rates are available until May 29.

See www.teragrid.org/tg09/ for updates.

The National Science Foundation will provide funding for 75 high school, undergraduate, and graduate students to participate in TeraGrid '09! [Read more about opportunities for student participation.](#)

TeraGrid '09 News

[Deadline for poster submissions extended to May 6.](#)
[Registration is now open for TeraGrid '09](#)
[TeraGrid '09 Keynote Speakers Include Ed Seidel, Paul Avery and Thomas Cheatham](#)

For Our Users

User Portal

Username: Password:

[What is the User Portal?](#)

Want to request a TeraGrid allocation? Need info on setting up or using a new account? Want to create a Science Gateway or publish a Data Collection? Check out the links below for a comprehensive introduction to TeraGrid resources.

- [XD Transition Information](#)
- [Data and Computation Resources](#)
- [Getting Started Guide](#)
- [User & System News](#)
- [Campus Champions Support Local Users](#)
- [Search the Knowledge Base](#)
- [Get a TeraGrid Allocation](#)
- [Email Helpdesk/Support](#)
- [Learn About Science Gateways](#)
- [Explore Advanced Support for TeraGrid Applications](#)

Science Highlights

Calendar of Events

3

Researcher requests access to SSG



Cornell University

Search Cornell

VirtualRDC News @ Cornell

Announcement

Preliminary program for NSF-Census-IRS Workshop on Synthetic Data and Confidentiality Protection posted

June 2nd, 2009

A preliminary but near-final program has been posted to <http://www.vrdc.cornell.edu/news/nsf-census-irs-workshop2009/>. Participants are still welcome to register. [\[More »\]](#)

Posted in [General](#) | [Edit](#) | [No Comments »](#)

SPSS readin files for QWI uploaded

June 18th, 2009

It took us a while, but we have uploaded instructions provided by a user on how to load `qwi` files in `spss`, including the relevant format file. [\[More »\]](#)

Posted in [Data](#) | [Edit](#) | [No Comments »](#)

Downtime for VRDC6401 later this week

June 16th, 2009

There will be a short downtime including reboot later this week (presumably on Friday morning) for VRDC6401. Your sessions will be terminated. Please plan accordingly.

Posted in [VirtualRDC @Cornell](#) | [Edit](#) | [No Comments »](#)

Subscribing to notification by email

June 10th, 2009

We have enabled a FeedBurner-moderated notification mechanism: you can now get email updates when a new post is published on this site. In order to subscribe, go to [Mailing lists](#). And [let us know](#) if you like this new mechanism.

Posted in [General](#) | [Edit](#) | [No Comments »](#)

Site search

Search for this text:

Search

Site Navigation

Front page

[open all](#) | [close all](#)

- [General Information](#)
- [Help for RDC proposal..](#)
- [Data @ VirtualRDC](#)
- [Available resources](#)
- [Classes and Tutorials](#)
- [Documentation](#)
- [NSF-Census-IRS Worksl](#)

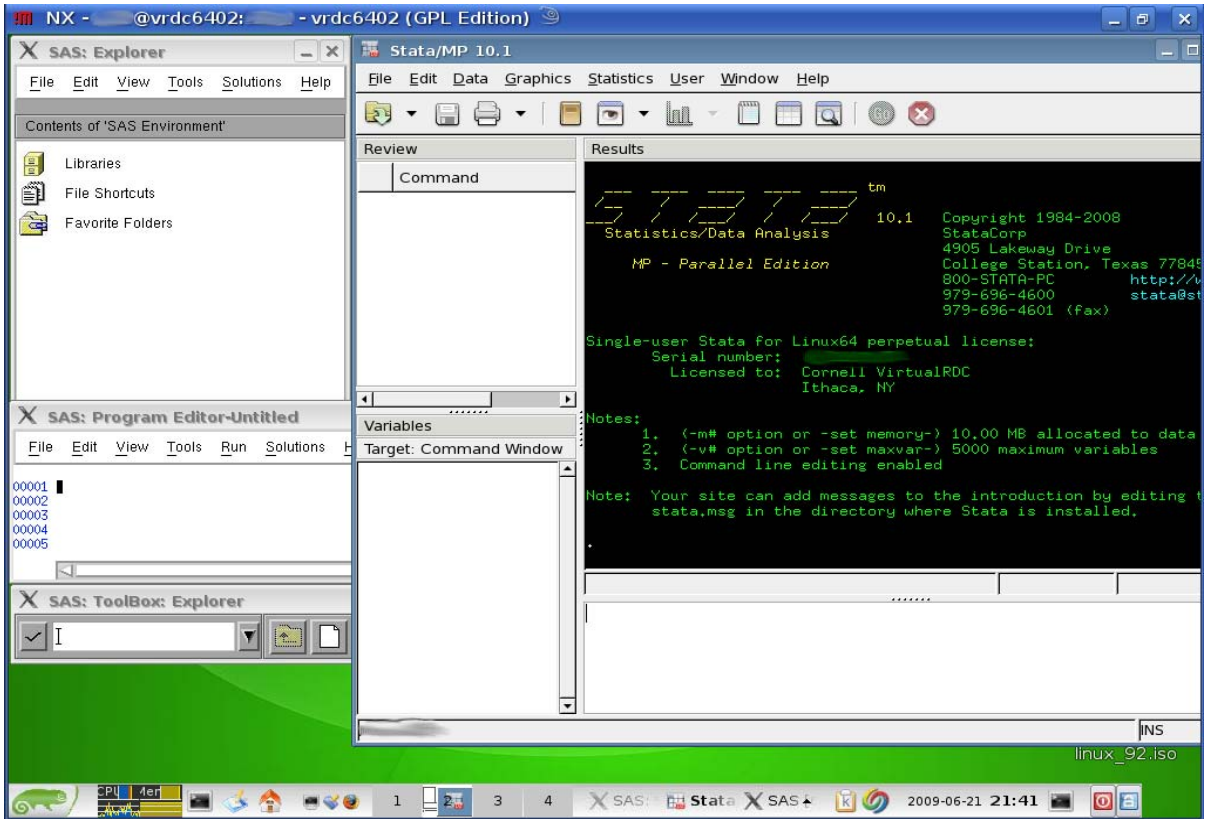
Recent articles

[open all](#) | [close all](#)

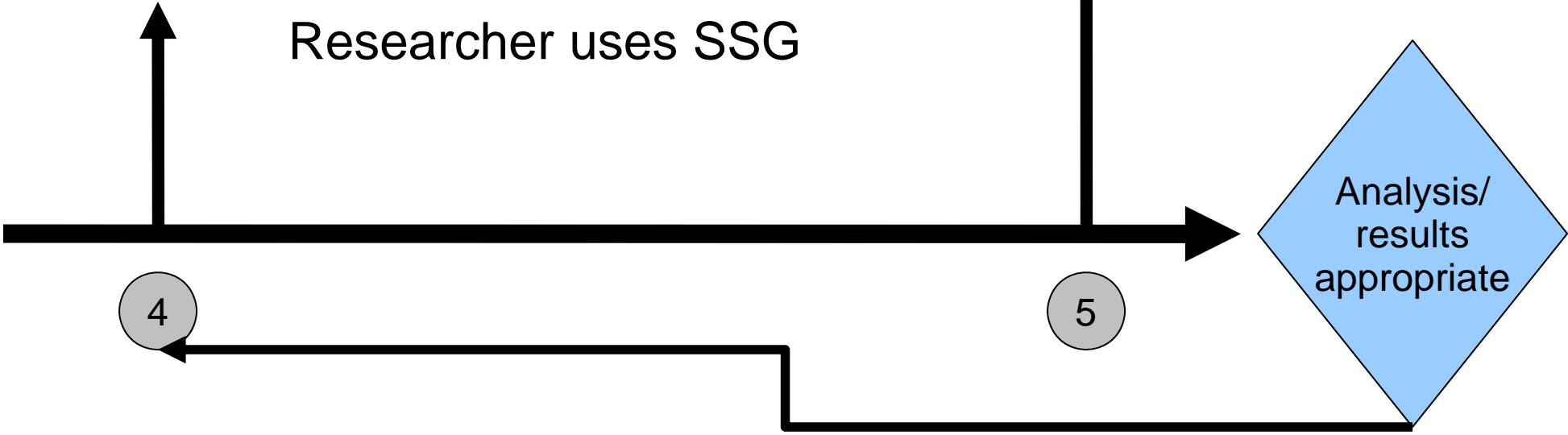
- [Data \(50\)](#)
- [Documents \(20\)](#)
- [Events \(84\)](#)
- [General \(63\)](#)
- [Hardware \(31\)](#)
- [NSF-Census-IRS Worksl](#)

Typical workflow (cont)

- (4) Data preparation and analysis on SSG
- (5) Full computation on TG
- Iterate (4) and (5) until satisfied
- End of project



Researcher uses SSG



4

Researcher accesses SSG

The screenshot displays a virtual machine interface with the following components:

- Operating System:** NX - @vrdc6402: - vrdc6402 (GPL Edition)
- SAS: Explorer:** Shows the file structure of the SAS environment, including Libraries, File Shortcuts, and Favorite Folders.
- SAS: Program Editor-Untitled:** A text editor window showing a list of lines (00001 to 00005).
- SAS: ToolBox: Explorer:** A toolbar with a checked checkbox and a dropdown menu.
- Stata/MP 10.1:** The main application window, showing the Stata logo and version information. The Results pane displays the following text:

```
tm
StataMP 10.1 Copyright 1984-2008
Statistics/Data Analysis StataCorp
MP - Parallel Edition 4905 Lakeway Drive
College Station, Texas 77845
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

Single-user Stata for Linux64 perpetual license:
Serial number: [REDACTED]
Licensed to: Cornell VirtualRDC
Ithaca, NY

Notes:
1. (-m# option or -set memory-) 10.00 MB allocated to data
2. (-v# option or -set maxvar-) 5000 maximum variables
3. Command line editing enabled

Note: Your site can add messages to the introduction by editing the
file stata.msg in the directory where Stata is installed.
```

The taskbar at the bottom shows the system date and time as 2009-06-21 21:41, along with icons for SAS, Stata, and other applications.

5

Researcher submits job on TG



TeraGrid™
User Portal

[Logout](#)

Welcome

Home | **My TeraGrid** | Resources | Documentation | Training | Consulting | Allocations

Systems Monitor | File Manager | HPC Queue Prediction | Remote Visualization | Science Gateways | Data Collections | User Responsibilities

TeraGrid Systems Monitor

[HPC Systems](#) | [Storage Systems](#) | [Advanced Vis Systems](#) | [Special Purpose Systems](#)

High Performance Systems

Name	Institution	System	Peak TFlops	Memory TBytes	Status	Load	Running Jobs	Queued Jobs	Other Jobs
Kraken	NICS	Cray XT5	608.00	129.00	Up	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	121	103	77
Ranger	TACC	Sun Constellation	579.40	123.00	Up	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	382	145	184
Abe	NCSA	Dell Intel 64 Linux Cluster	89.47	9.38	Up	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	89	495	374
Lonestar	TACC	Dell PowerEdge Linux Cluster	62.16	11.60	Up	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	136	281	0
Steele	Purdue	Dell Intel 64 Linux Cluster	60.00	12.40	Up*	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	386	309	89
Queen Bee	LONI	Dell Intel 64 Linux Cluster	50.70	5.31	Up	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	113	205	35
Lincoln	NCSA	Dell/Intel PowerEdge 1950	47.50	3.00	Up	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	1	2	0
Big Red	IU	IBM e1350	30.60	6.00	Up*	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	885	1983	19
BigBen	PSC	Cray XT3	21.50	4.04	Up	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	16	139	50
Mercury	NCSA	IBM Itanium2 Cluster	10.23	4.47	Up	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	123	598	0
Cobalt	NCSA	SGI Altix	6.55	3.00	Up	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	33	263	112
Frost	NCAR	IBM BlueGene/L	5.73	0.51	Up	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	1	25	0
Pope	PSC	SGI Altix 4700	5.00	1.54	Up*	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	19	70	31
TeraGrid Cluster	SDSC	IBM Itanium2 Cluster	3.10	1.02	Up	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	7	0	1
TeraGrid Cluster	UC/ANL	IBM Itanium2 Cluster	0.61	0.24	Up*	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	0	0	1
NSTG	ORNL	IBM IA-32 Cluster	0.34	0.07	Up	<div style="width: 100%; height: 10px; background-color: #0070c0;"></div>	4	0	0
Total:			1580.89	314.58			2316	4618	973

Researcher...

- ...transfers data to TG (gridftp)
- ...submits jobs on TG
- ...iterates back and forth until project is completed

Other Cornell resources

- Cornell Matlab cluster (same 10Gbps network)
- Additional storage if needed by project (rental on a yearly basis)

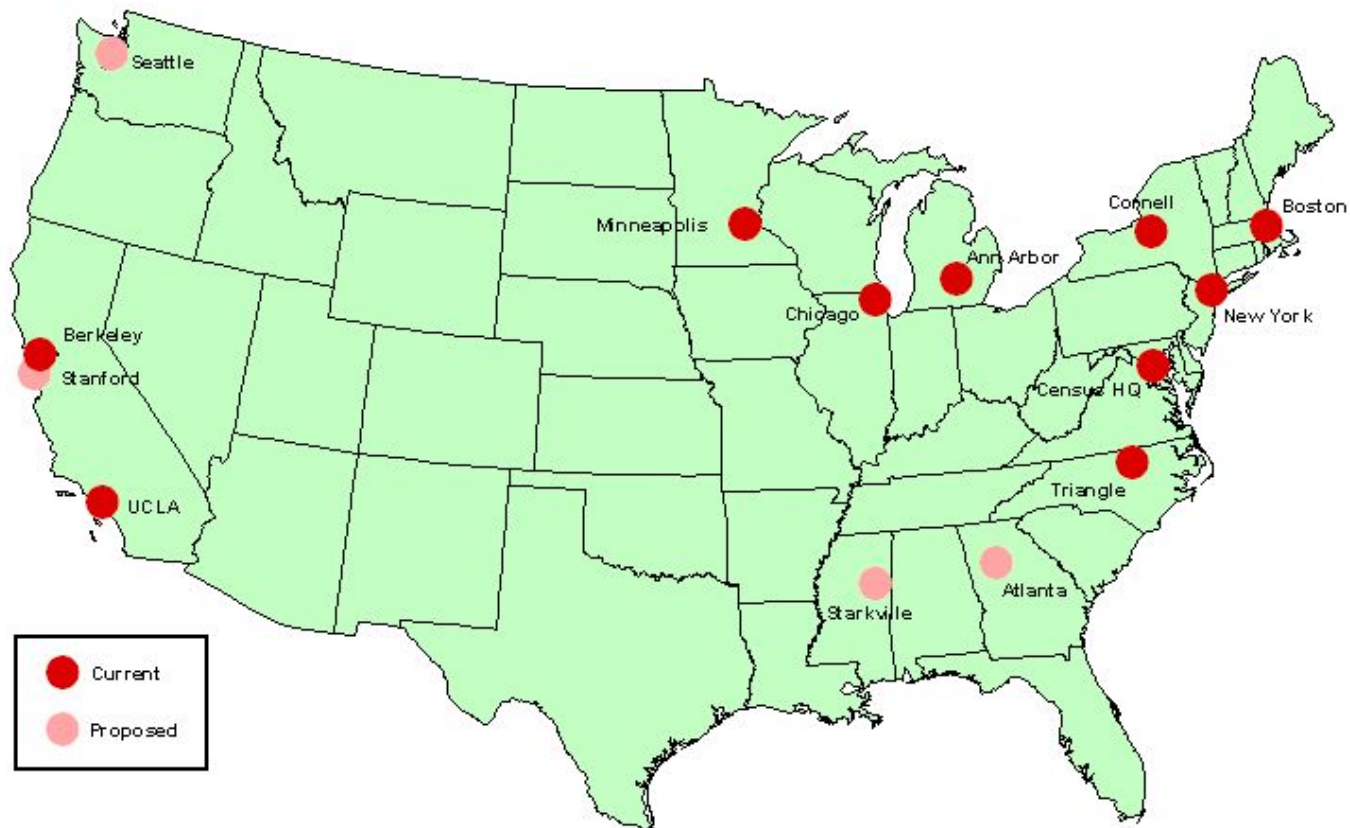
VirtualRDC @ Cornell University

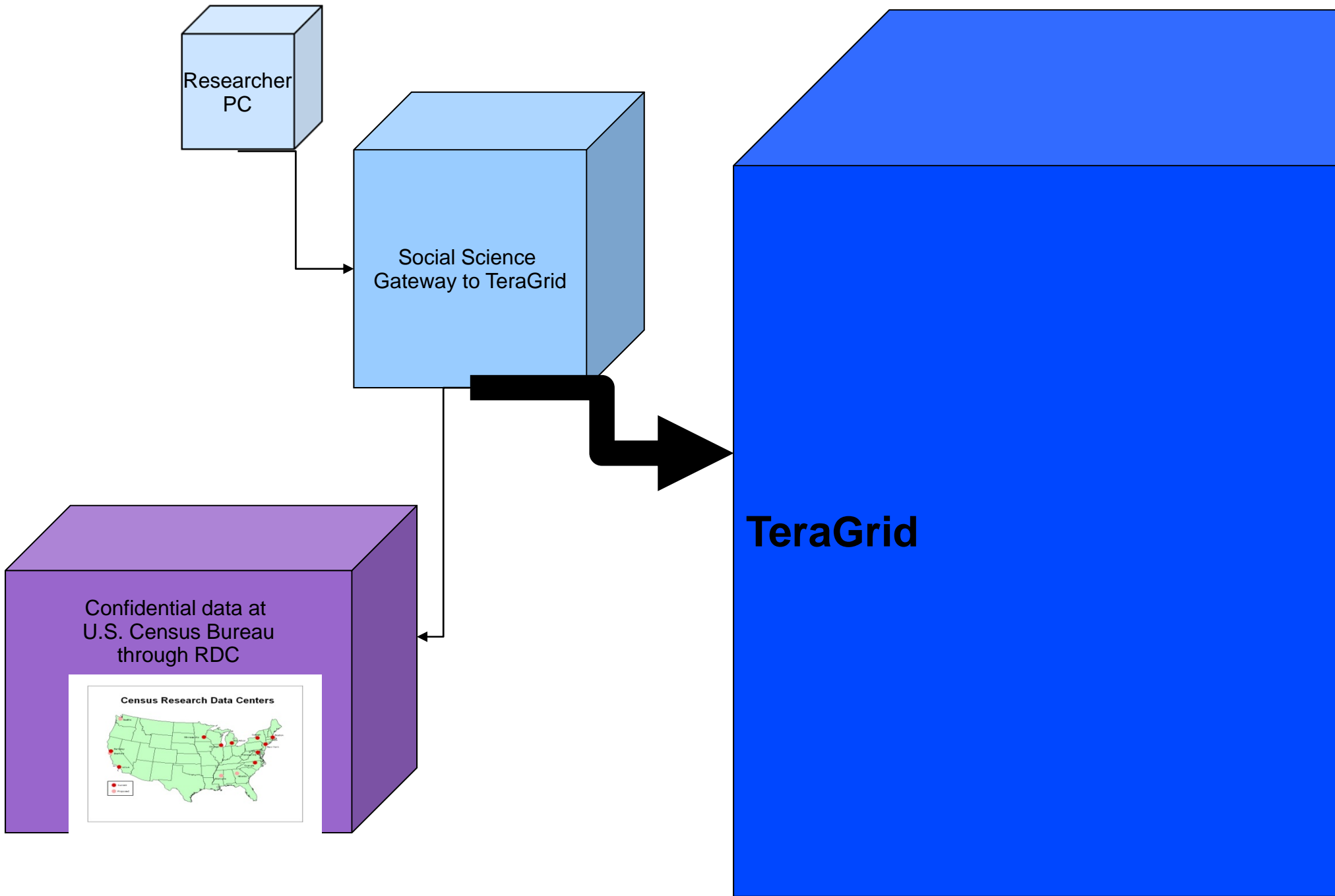
(also known as Social Science Gateway
to Teragrid)

VirtualRDC provides:

- Environment replicating the RDC compute center (Linux GUI, directory structure, etc.)
- Zero-obs datasets allowing some basic programming preparation
- Stepping-stone to projects in the Census RDC

Census Research Data Centers





Social Science Gateway @Cornell

Funding through NSF (pending)

Available 4th quarter 2009
will be announced at
<http://www.vrdc.cornell.edu>

Contact:

Lars Vilhuber (lars.vilhuber@cornell.edu)

Or virtualrdc@cornell.edu

TODO

- Startup...
- Update website
- Tutorial on workflow, tools
- Integrate GUI tools for GridFTP transfers
 - Grifi
 - Globus Java Gui