SciServer

Our work on Databases and SciServer

Gerard Lemson, Alex Szalay
IDIES, Johns Hopkins University

sciserver.org

JOHNS HOPKINS UNIVERSITY

NSF

idies

Data dissemination with Relational Databases

- Sloan Digital Sky Survey (SDSS)
- *Millennium* Cosmological Simulations
  - Time evolution in “merger trees”, 3D spatial queries
  - Custom SQL, MyDBs

Collaborative data-driven science

SciServer

Image Credit: Sloan Digital Sky Survey

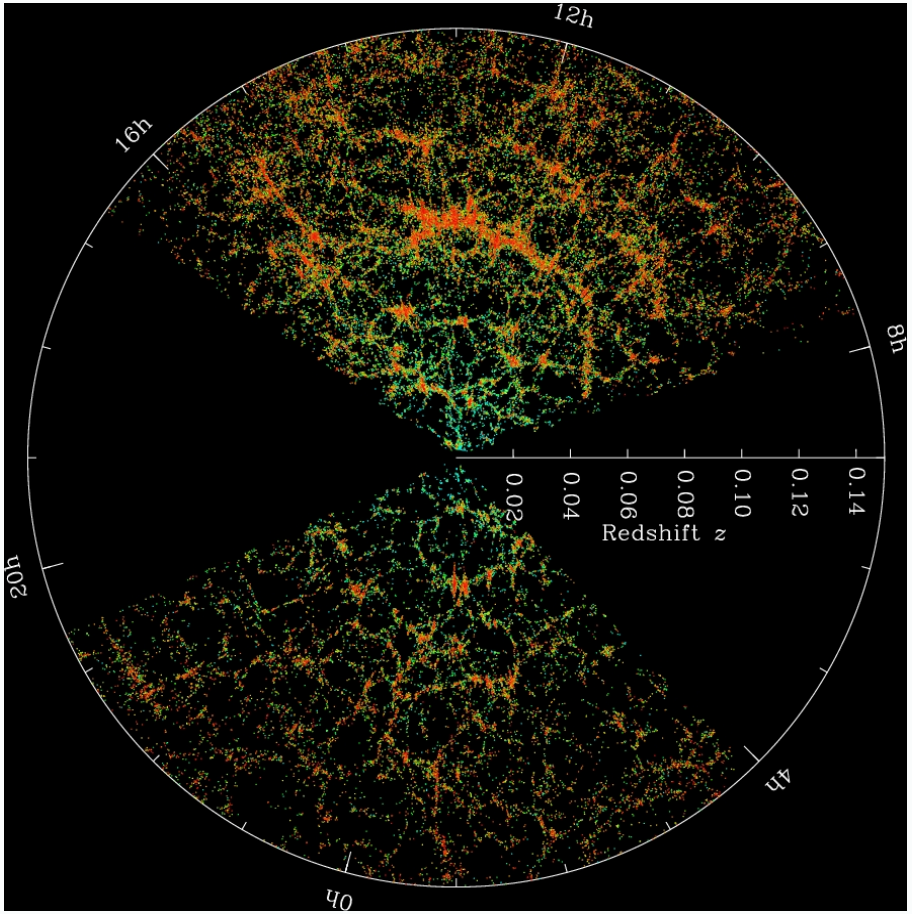
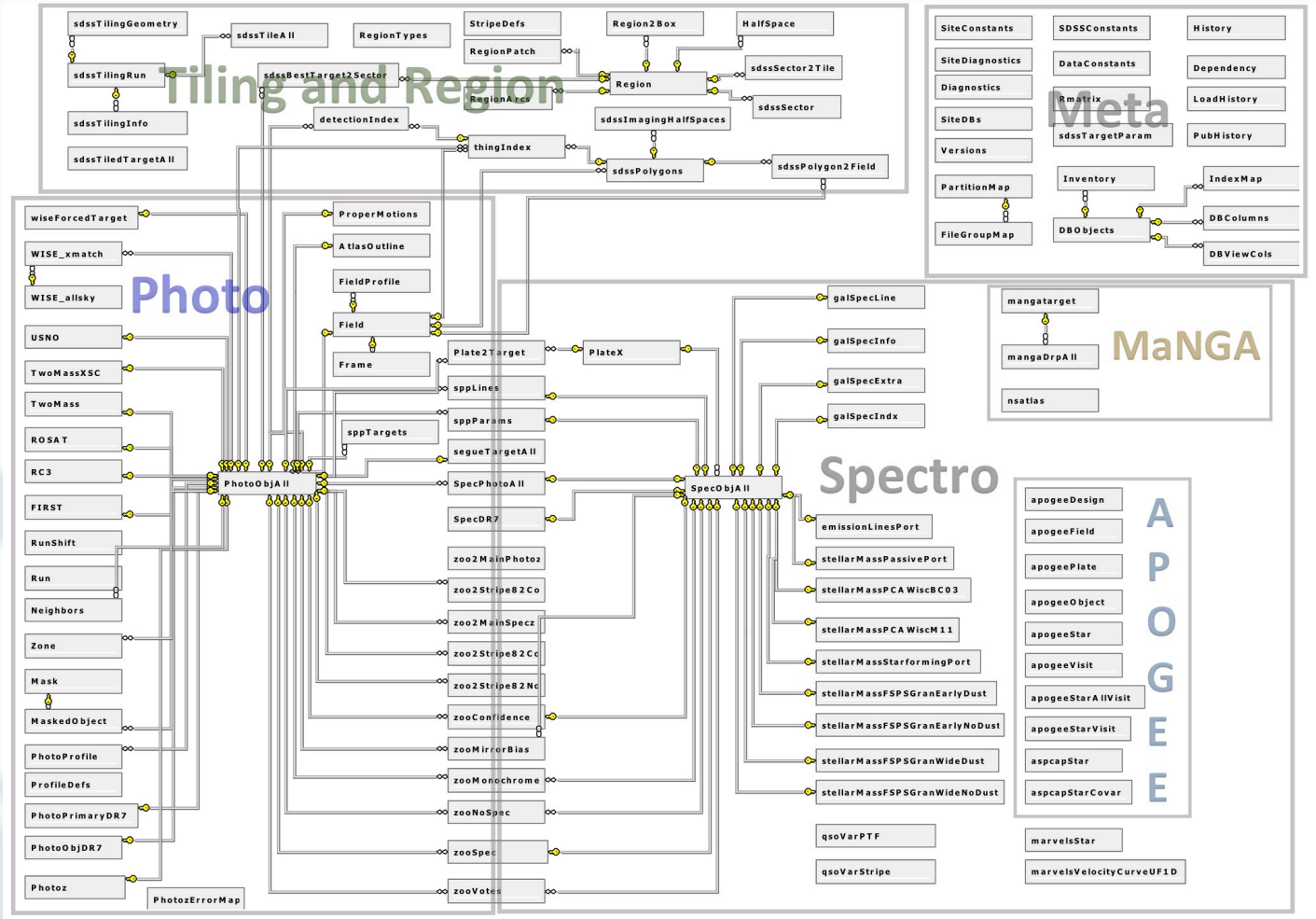


Image Credit: M. Blanton and SDSS

Collaborative data-driven science

SciServer



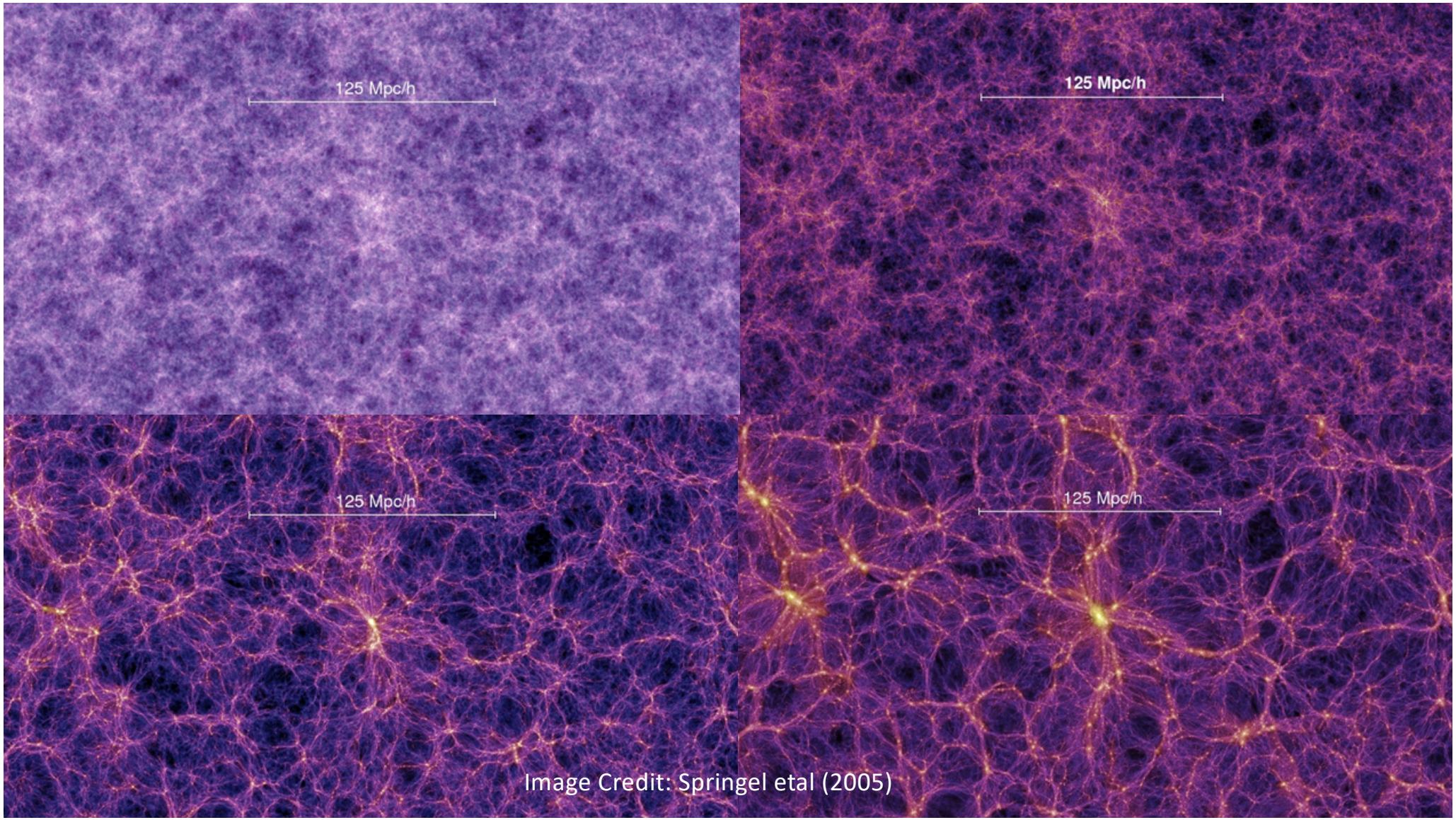
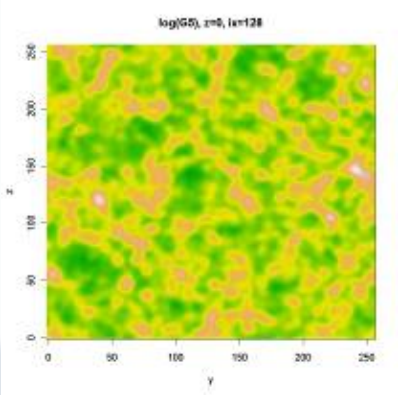
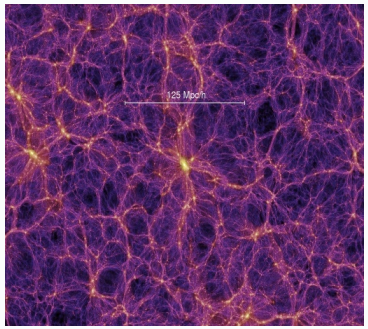


Image Credit: Springel etal (2005)

Simulation post-processing pipeline

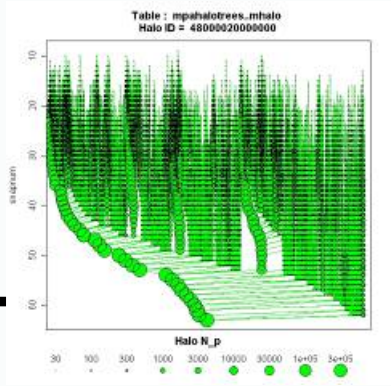


Gridding

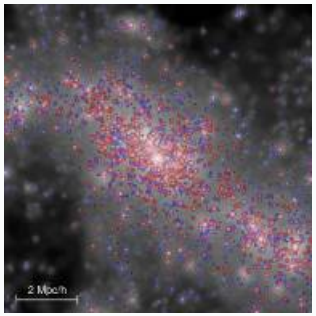


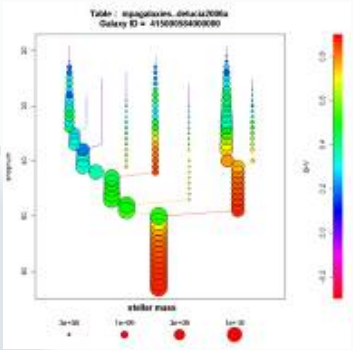
Raw particle data

Cluster finders

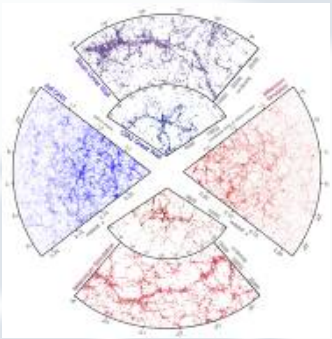


Simulating galaxy formation



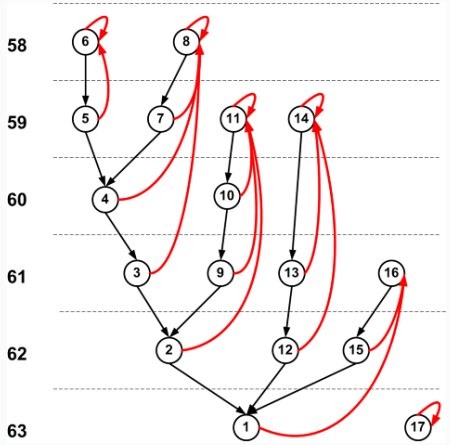


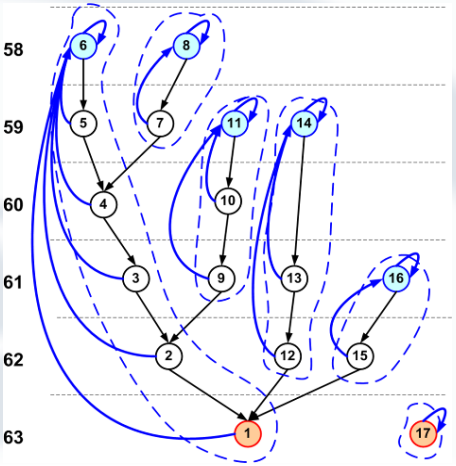
Virtual observations

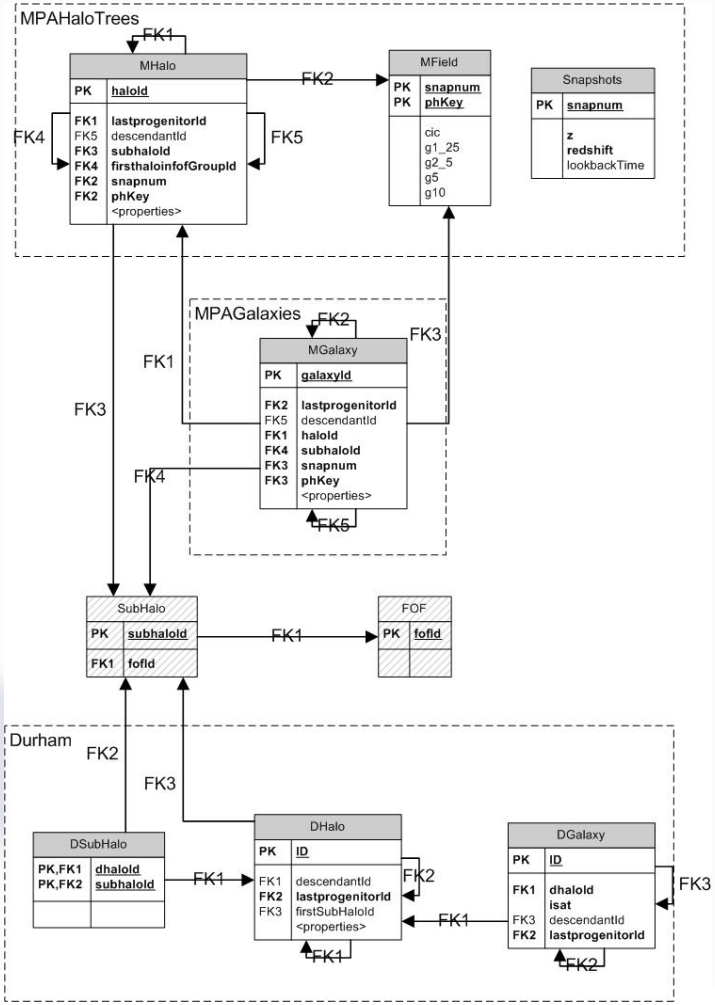


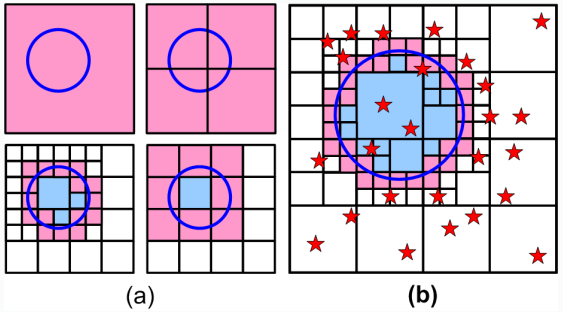
Collaborative data-driven science

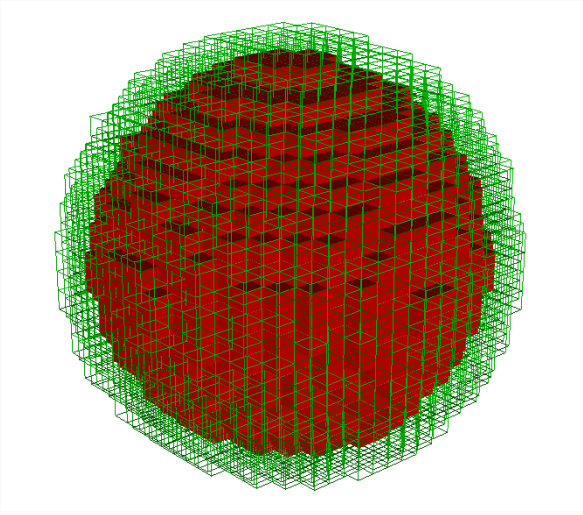
SciServer





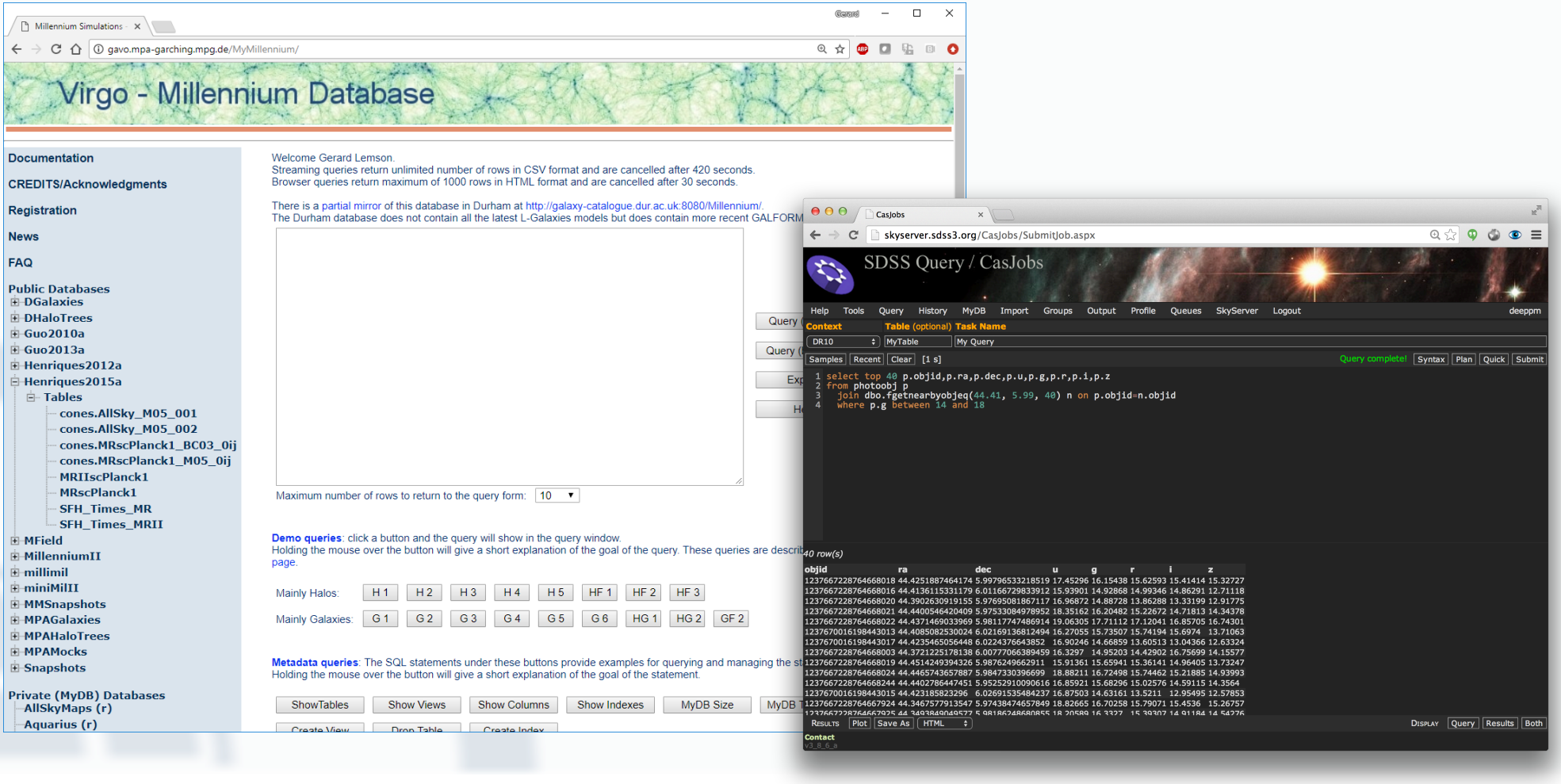






Collaborative data-driven science

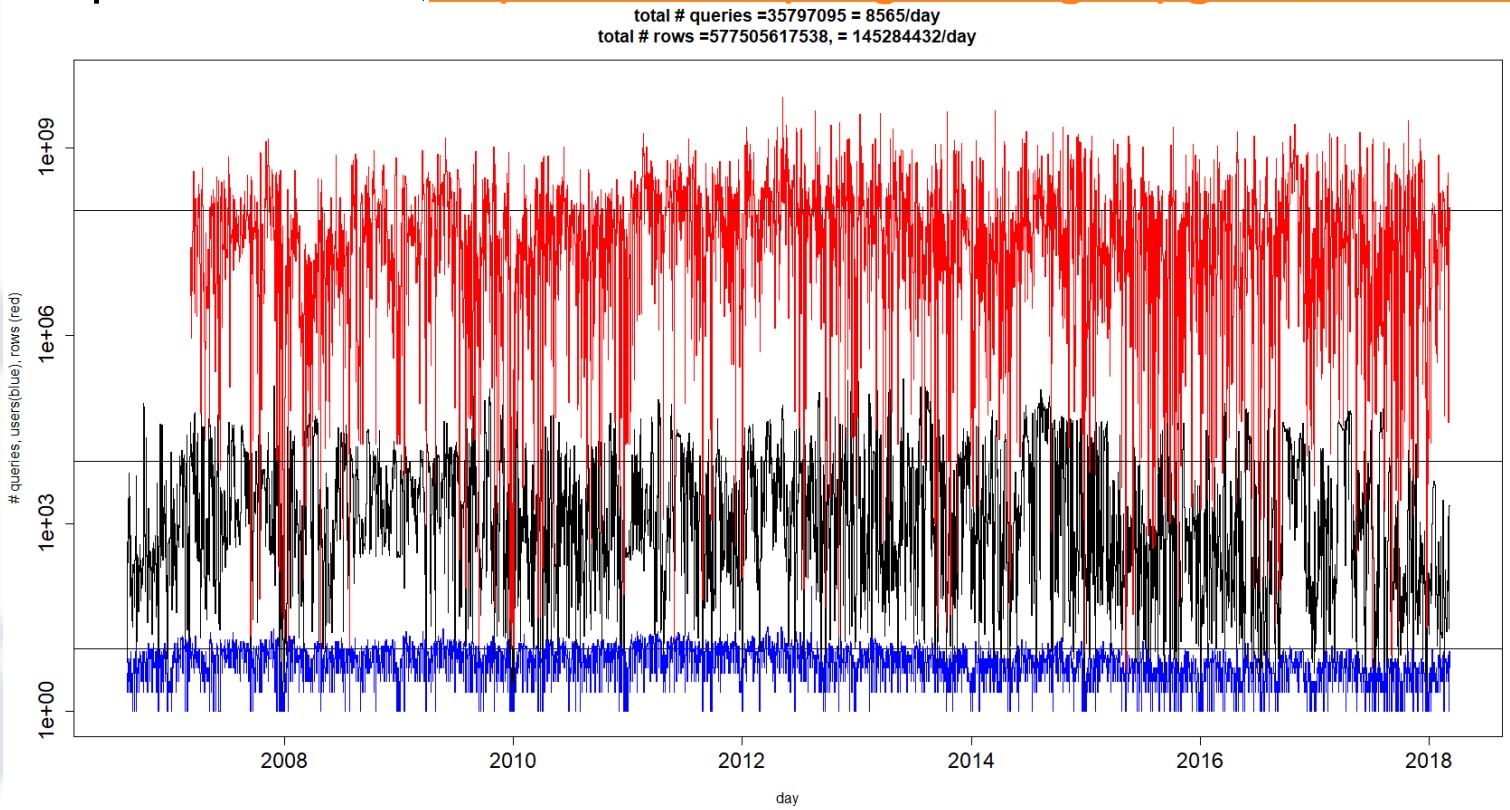
SciServer



Access statistics

>1000 registered users

>1000 publications (http://www.mpa-garching.mpg.de/millennium/)

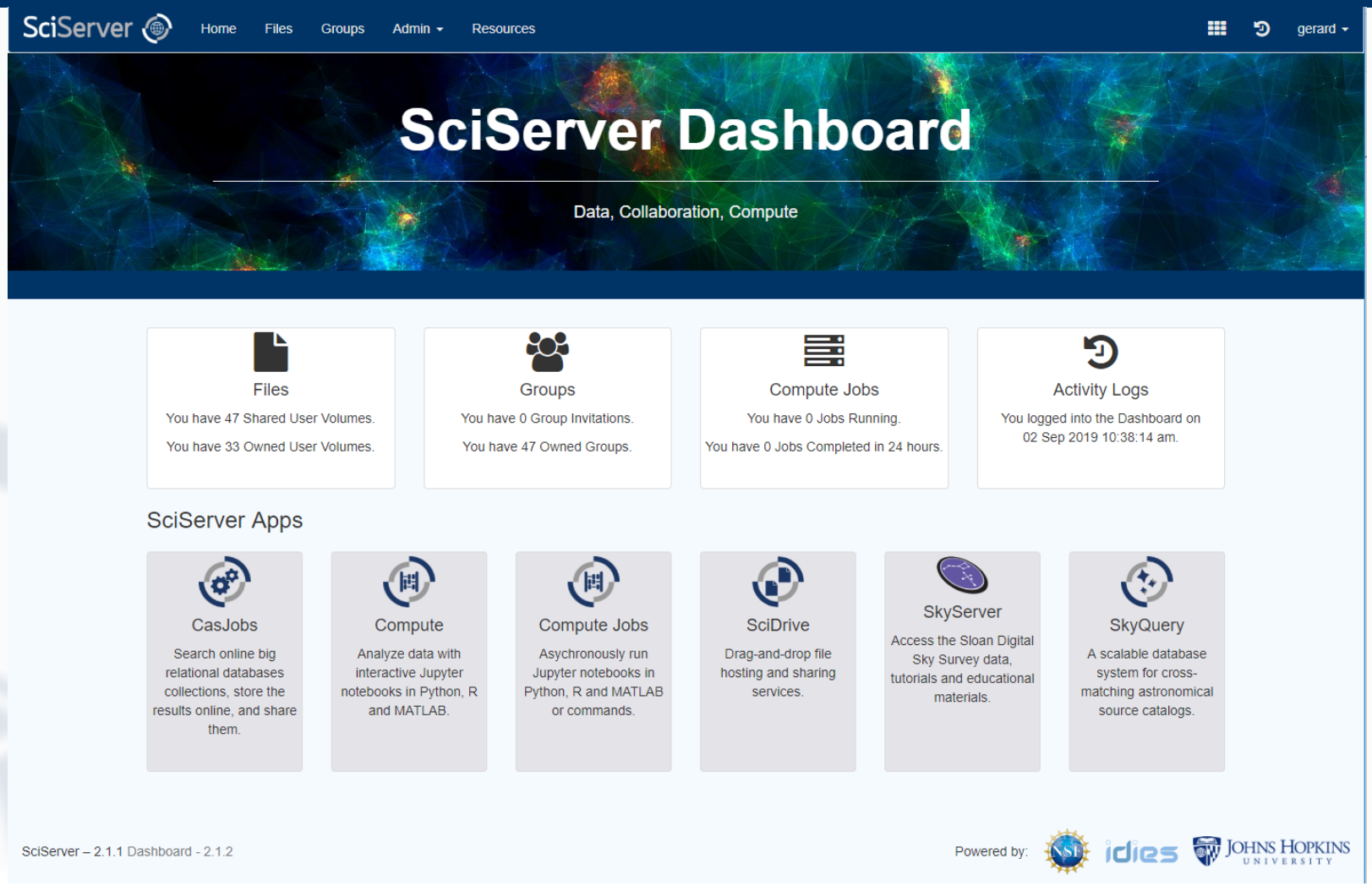


SciServer: beyond databases and astrophysics

- **SciServer** is a system allowing Science Researchers across multiple domains to host and share their datasets and provide query and analysis tools for collaborative research.
- **Core Services:**
  - Science Data Hosting (Files and Databases)
  - Personal Storage (Files and Database)
  - Query hosted databases (Custom SQL)
  - Computational analysis on hosted data (Python R, Matlab, custom apps)
  - Collaboration and Sharing (create user groups, flexible access controls)
  - API Integration (easy to wrap SciServer with custom UIs)

Collaborative data-driven science

SciServer



Linking relational catalog to image archives

jupyter 2. Stripe82-coadd-Copy1 Last Checkpoint: 3 minutes ago (unsaved changes)

File Edit View Insert Cell Kernel Help | Python 3

Code | CellToolbar

```
import SciServer.LoginPortal as Login
token=Login.getToken()
import pandas
import tables
import numpy as np
import astropy
from astropy.io import fits
from astropy import wcs
import skimage.io
import urllib
import os
import matplotlib
import matplotlib.pyplot as plt
```

In [18]:

```
sql="""
SELECT a.objid as head, c.objid2 as match, b.matchcount,
       p.fieldid as head_field, d.fieldid as match_field,
       dbo.fGetUrlFitsCFrame(d.fieldid, 'g') as fits_g,
       dbo.fGetUrlFitsCFrame(d.fieldid, 'r') as fits_r,
       dbo.fGetUrlFitsCFrame(d.fieldid, 'z') as fits_z,
       p.ra, d.ra as match_ra, p.dec, d.dec as match_dec
       , p.petror90_r
  from (select top 1 * from galaxy where objId=865819437896092809) a
   join matchhead b on a.objid=b.objid        -- join with matchhead
   join photoobj p on a.objid=p.objid         -- get matchhead photoobj
   join match c on c.objid1=b.objid           -- join with all the matches
   join photoobjall d on c.objid2=d.objid     -- get match photoobj
order by d.fieldid                              -- order by matchhead objid
"""
queryResponse = SciServer.CasJobs.executeQuery(sql, "Stripe82",token=token)
obss = pandas.read_csv(queryResponse,index_col=None)
obss[:10]
```

Out[18]:

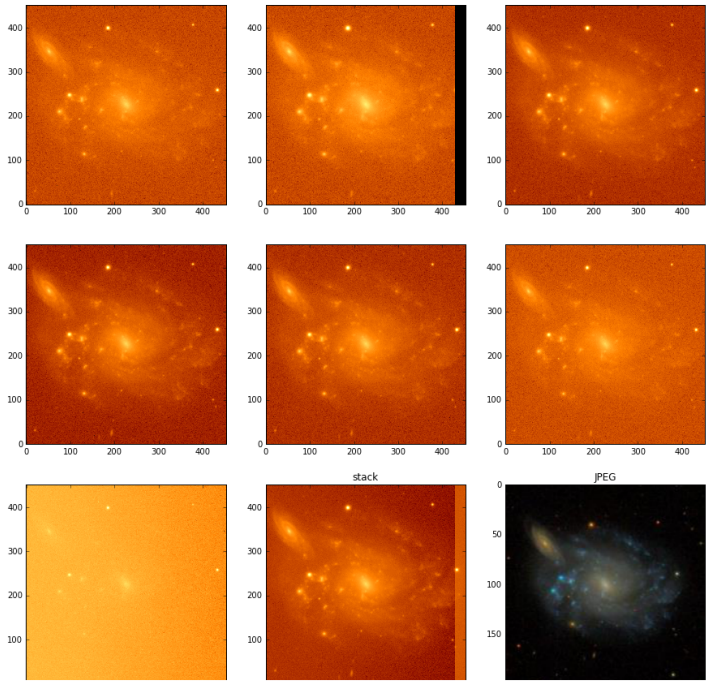
	head	match	matchcount	head_field	match_field	fits_g
0	865819437896092809	865819443049955320	57	865819437896092768	865819443049955328	http://das.sdss.org/imaging/5622/40/corr/
1	865819437896092809	865819447774294837	57	865819437896092768	865819447774294835	http://das.sdss.org/imaging/5633/40/corr/
2	865819437896092809	865819451637537182	57	865819437896092768	865819451637537776	http://das.sdss.org/imaging/5642/40/corr/
3	865819437896092809	865819458508351079	57	865819437896092768	865819458508351078	http://das.sdss.org/imaging/5658/40/corr/
4	865819437896092809	865819480416301877	57	865819437896092768	865819480416301875	http://das.sdss.org/imaging/5709/40/corr/
5	865819437896092809	865819495447075229	57	865819437896092768	865819495447075225	http://das.sdss.org/imaging/5744/40/corr/
6	865819437896092809	865819501890791016	57	865819437896092768	865819501890791014	http://das.sdss.org/imaging/5759/40/corr/
7	865819437896092809	865819504465104080	57	865819437896092768	865819504465104076	http://das.sdss.org/imaging/5765/40/corr/
8	865819437896092809	865819506615123972	57	865819437896092768	865819506615123968	http://das.sdss.org/imaging/5770/40/corr/
9	865819437896092809	865819511339574889	57	865819437896092768	865819511339574886	http://das.sdss.org/imaging/5781/40/corr/

jupyter 2. Stripe82-coadd-Copy1 Last Checkpoint: a minute ago (autosaved)

File Edit View Insert Cell Kernel Help | Python 3

Code | CellToolbar

Out[17]: <matplotlib.text.Text at 0x7f2aa174a6d8>

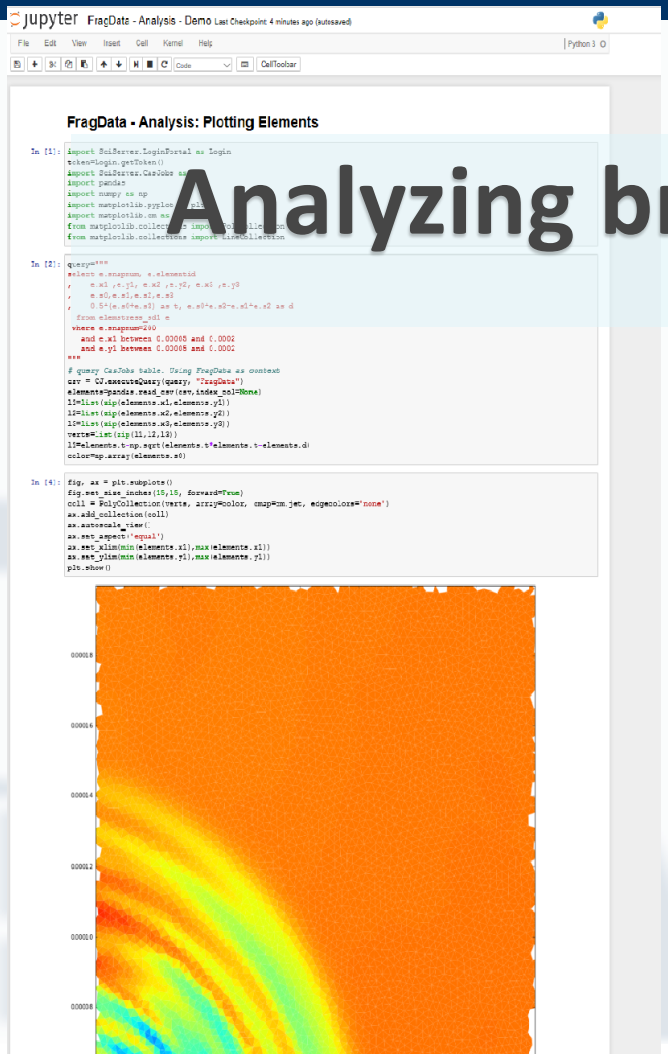


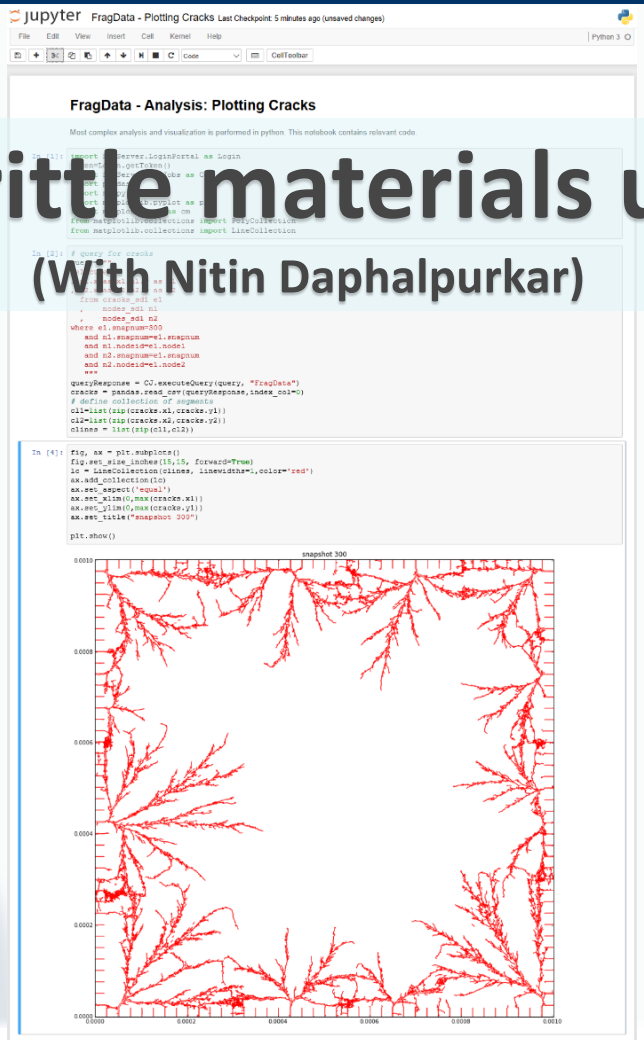
Collaborative data-driven science

SciServer

Analyzing brittle materials under stress

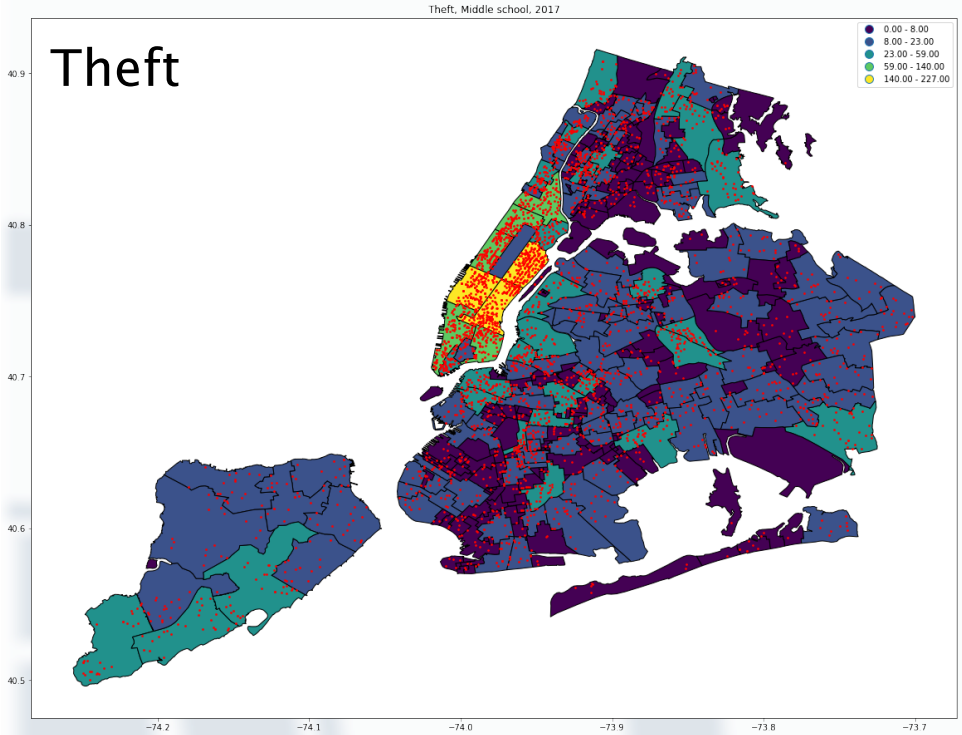
(With Nitin Daphalpurkar)

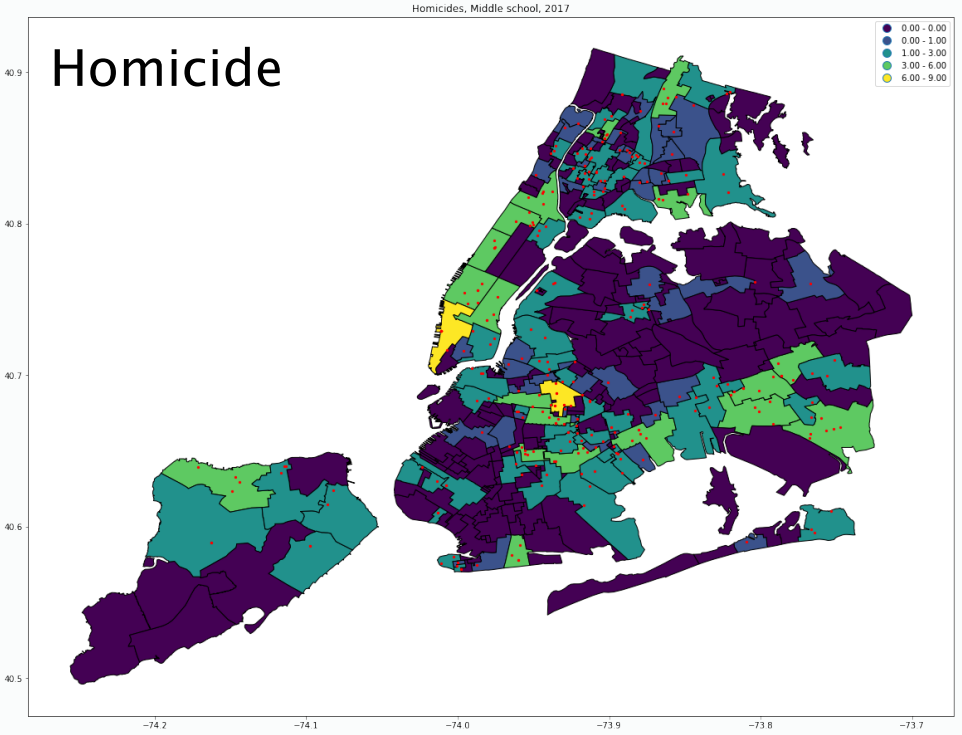




Crimes in school districts in NY.
To be correlated to school attendance

with J. Burdick-Will





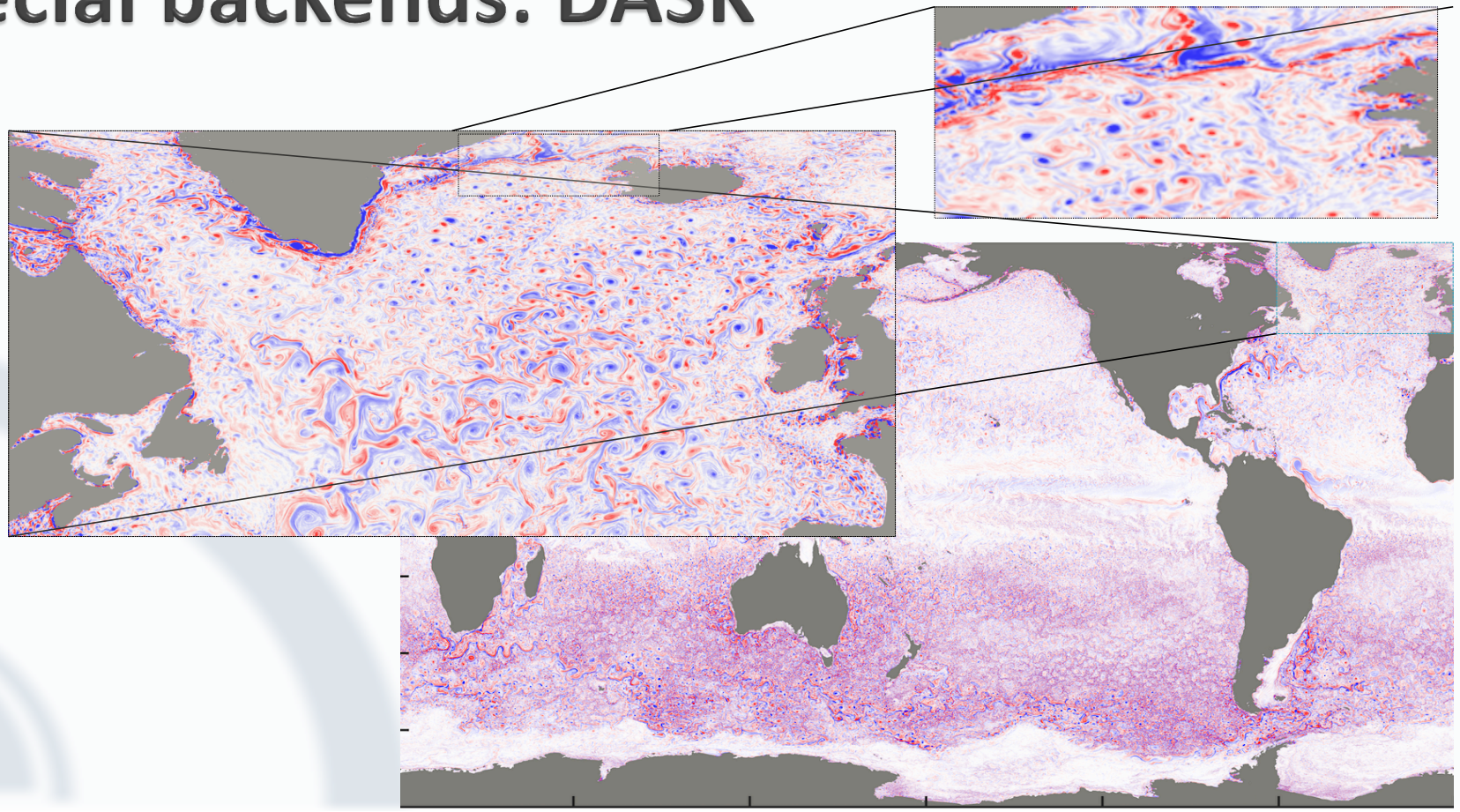
Analyzing metre in middle English poetry

(Christopher Cannon, Mark Patton)

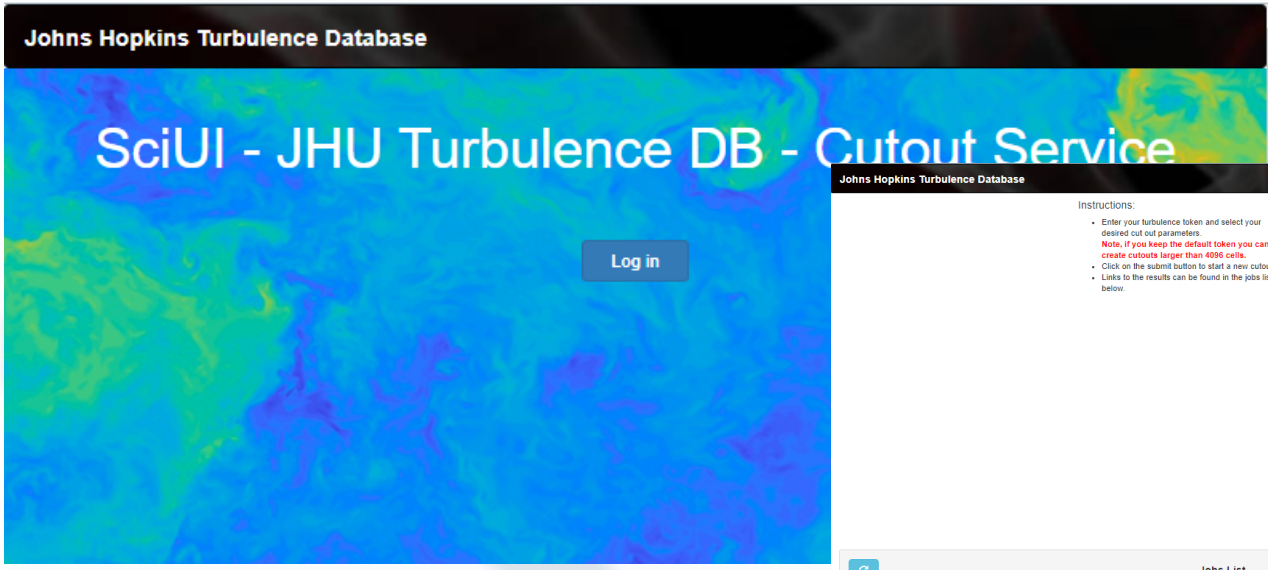
In [6]:
```
# iterate over features and print the original and best_parse version of each line
for line in txt.lines():
    best_parse = line.bestParse()  # most plausible parse
    all_parses = line.allParses()  # all plausible parses

    first_word = line.words()[0]
    last_syllable = line.syllables()[-1]
    last_syllable_rime = line.rimes()[-1]
    last_syllable_rime_phonemes = last_syllable_rime.phonemes()

    # print original line and best parsed version nicely, assuming maximum linelength< MAXLEN=60
    MAXLINE=60
    sline=str(line)
    blanks=" "*(MAXLINE-len(sline))
    print sline,blanks,best_parse
```

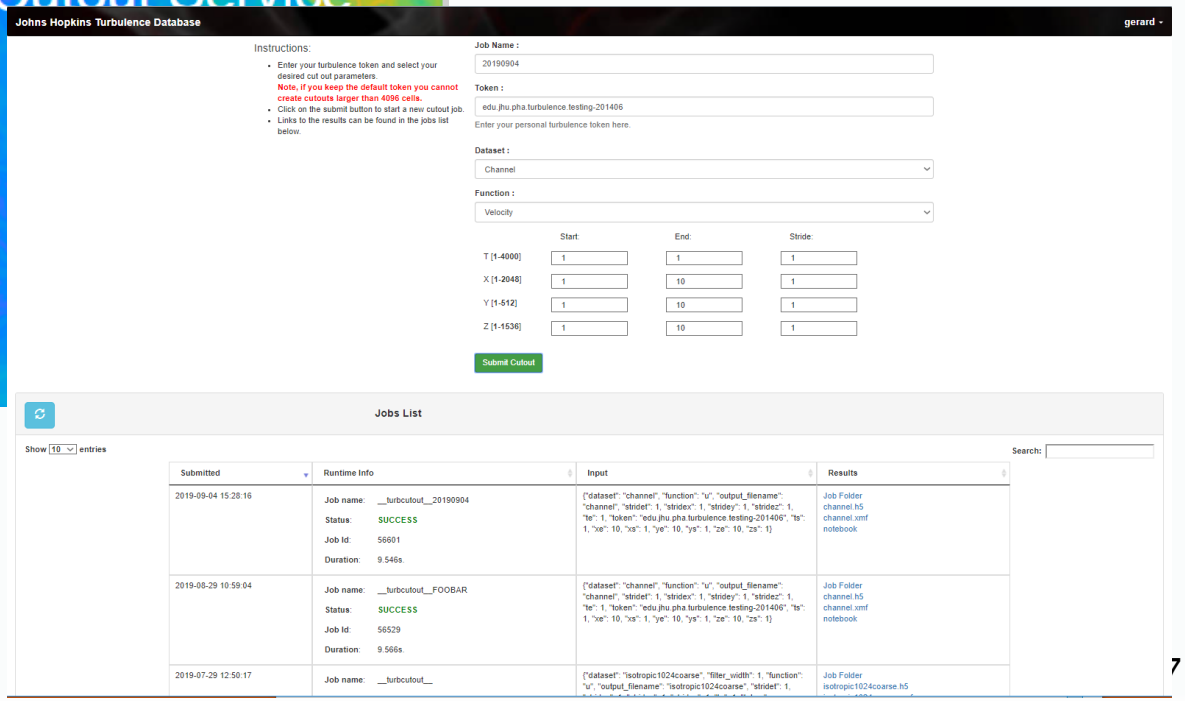
```
Turning and turning in the widening gyre                    TURN|ing.and*|TURN|ing|IN|the|WIDEN|ing|GY|re
The falcon cannot hear the falconer;                        the|FAL|con|CAN|not|HEAR|the|FAL|coner
Things fall apart; the centre cannot hold;                  things*|FALL|ap|ART|the|CEN|tre|CAN|not|HOLD
Mere anarchy is loosed upon the world,                      mere*|AN|ar|CHY*|is|LOOSED|up|ON|the|WORLD
The blood- dimmed tide is loosed, and everywhere            the|BLOOD|dimmed*|TIDE|is|LOOSED|and|EV|ery|WHERE
The ceremony of innocence is drowned;                       the|CER|e|MO|ny.of*|IN|no|CENCE*|is|DROWNED
The best lack all conviction, while the worst               the|BEST|lack*|ALL|con|VIC|tion|WHILE|the|WORST
Are full of passionate intensity.                           are|FULL|of|PAS|sion|ATE*|in|TEN|si|TY*
```

Special backends: DASK



Custom user interfaces





Collaborative data-driven science

SciServer