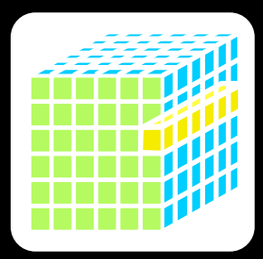
Seamless Data Sharing & High-Dimensional Visualization in Astronomy (and Beyond)



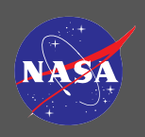
Alyssa A. Goodman
Harvard Smithsonian Center for Astrophysics
Radcliffe Institute for Advanced Study
DSI Steering Committee Member
@aagie

with A. Accomazzi, A. Beers, C. Beaumont, C. Borgman, M. Borkin, H. Chen, M. Crosas, J. Fay, D. Hogg, J. Kalpathy-Cramer, M. Kurtz, D. Lang, A. Muench, J. Peek, A. Pepe, P. Qian, T. Robitaille, B. Rosen, P. Rosenfield, J. Steffen, L. Trouille, P. Udomprasert, C. Zucker, C. Wong, and many more

HARVARD UNIVERSITY

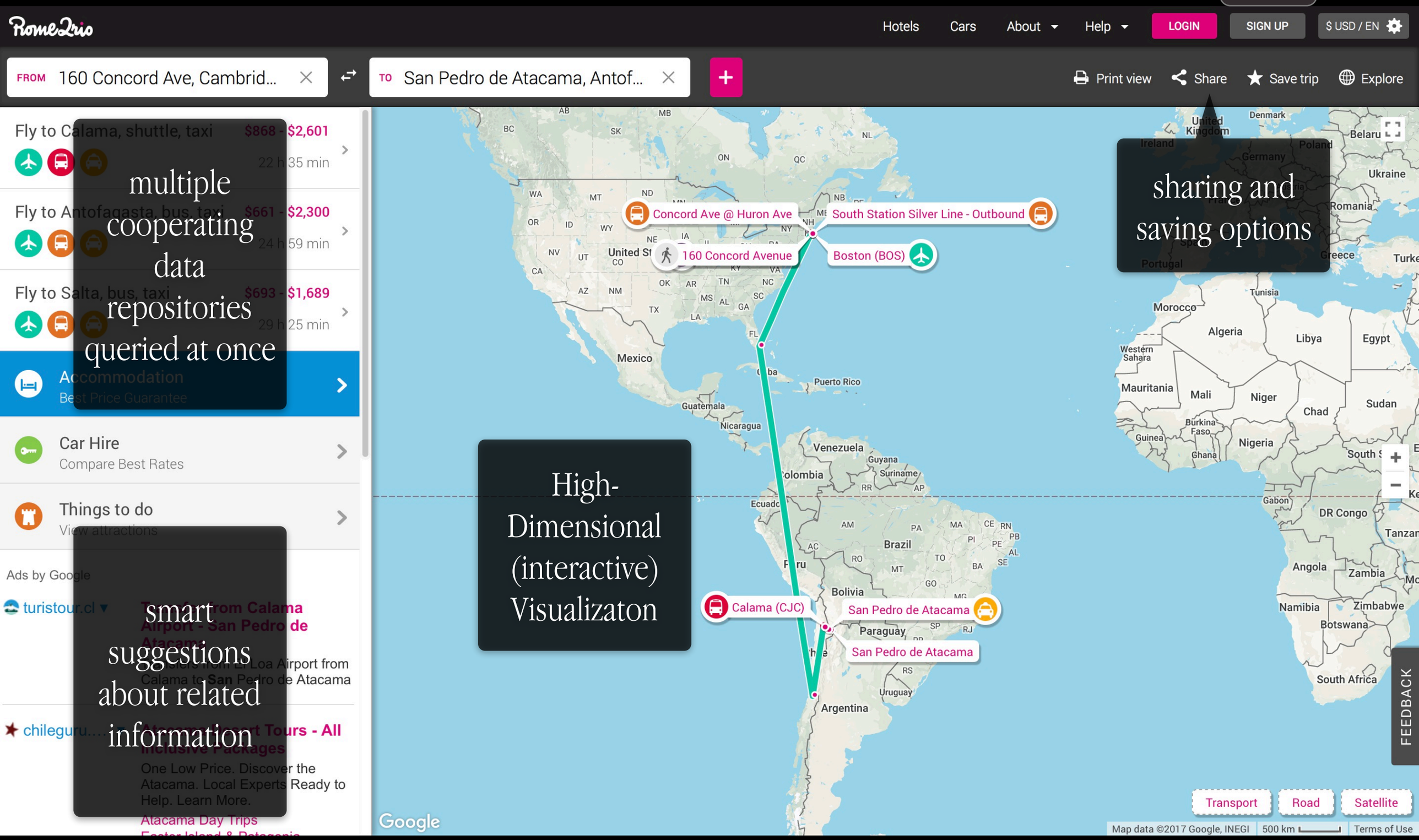
Microsoft Research

Authorea



AAS AMERICAN ASTRONOMICAL SOCIETY
Enhancing and sharing humanity's scientific understanding of the universe since 1899.

Let's begin with "beyond."



multiple cooperating data repositories queried at once

sharing and saving options

High-Dimensional (interactive) Visualizaton

smart suggestions about related information

https://www.rome2rio.com/s/160-Concord-Ave-Cambridge-MA-02138-USA/San-Pedro-de-Atacama-Antofagasta-Region-Chile

And now, Astronomy.

WorldWide Telescope

Explore | Guided Tours | Search | Communities | View | Settings | Install Windows Client | Sign Out

Collections >

1 of 2

New VAMP Feeds | Constellations | Solar System (Sky) | All-Sky Surveys | Spitzer Studies | Chandra Studies | Hubble Studies | Astrophotography | Radio Studies | NOAO Studies | Gemini Studies | Messier Catalog

sharing and saving options

multiple cooperating data repositories queried at once

High-Dimensional (interactive) Visualizaton

Look At: Sky

Imagery: Digitized Sky Survey (Color)

Image Crossfade

Tracking Andromeda Galaxy

1 of 2

N

Andromeda 01:44:01

RA: 00h42m44s
Dec: +41:16:09

Three Faces of | Young and Old | Andromeda Galaxy | Andromeda Galaxy | Andromeda Galaxy | Andromeda Galaxy | M31 | M31 | M31 | Andromeda Galaxy

smart suggestions about related information

http://worldwidetelescope.org/webclient/?loggedIn=true#/place=0.389&cf=77&ra=0.71231&dec=41.26917&fov=1.73327

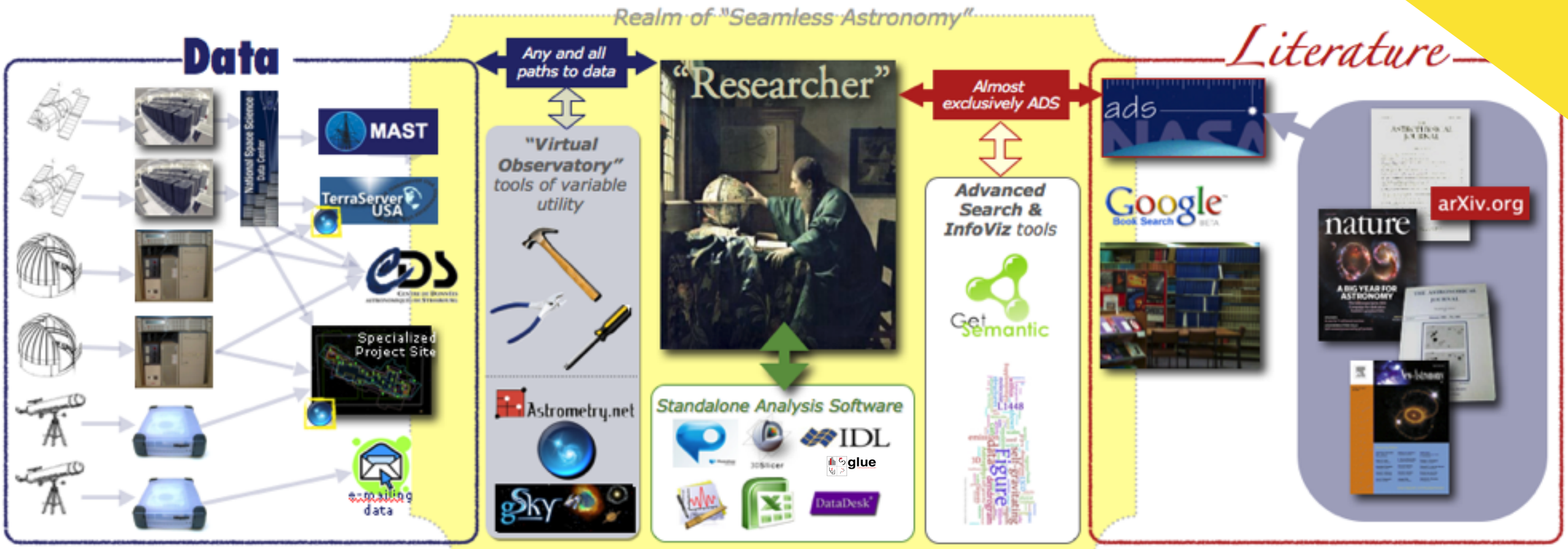
And now, Astronomy.

http://worldwidetelescope.org/webclient/?loggedIn=true#/place=0.389&cf=77&ra=0.71231&dec=41.26917&fov=1.73327

SEAMLESS ASTRONOMY

Linking scientific data, publications, and communities

2010



projects.iq.harvard.edu/seamlessastronomy

Smithsonian

Microsoft Research

SEAMLESS ASTRONOMY

Linking scientific data, publications, and communities

2017

astrophysics data system

Authorea

The Dataverse Network™ Project

Astrometry.net

flickr

ZOONIVERSE
REAL SCIENCE ONLINE

glue
multidimensional data exploration

ADS ALL SKY SURVEY

edX

WorldWide Telescope
Ambassadors Program

WIKIPEDIA
The Free Encyclopedia

ALADIN

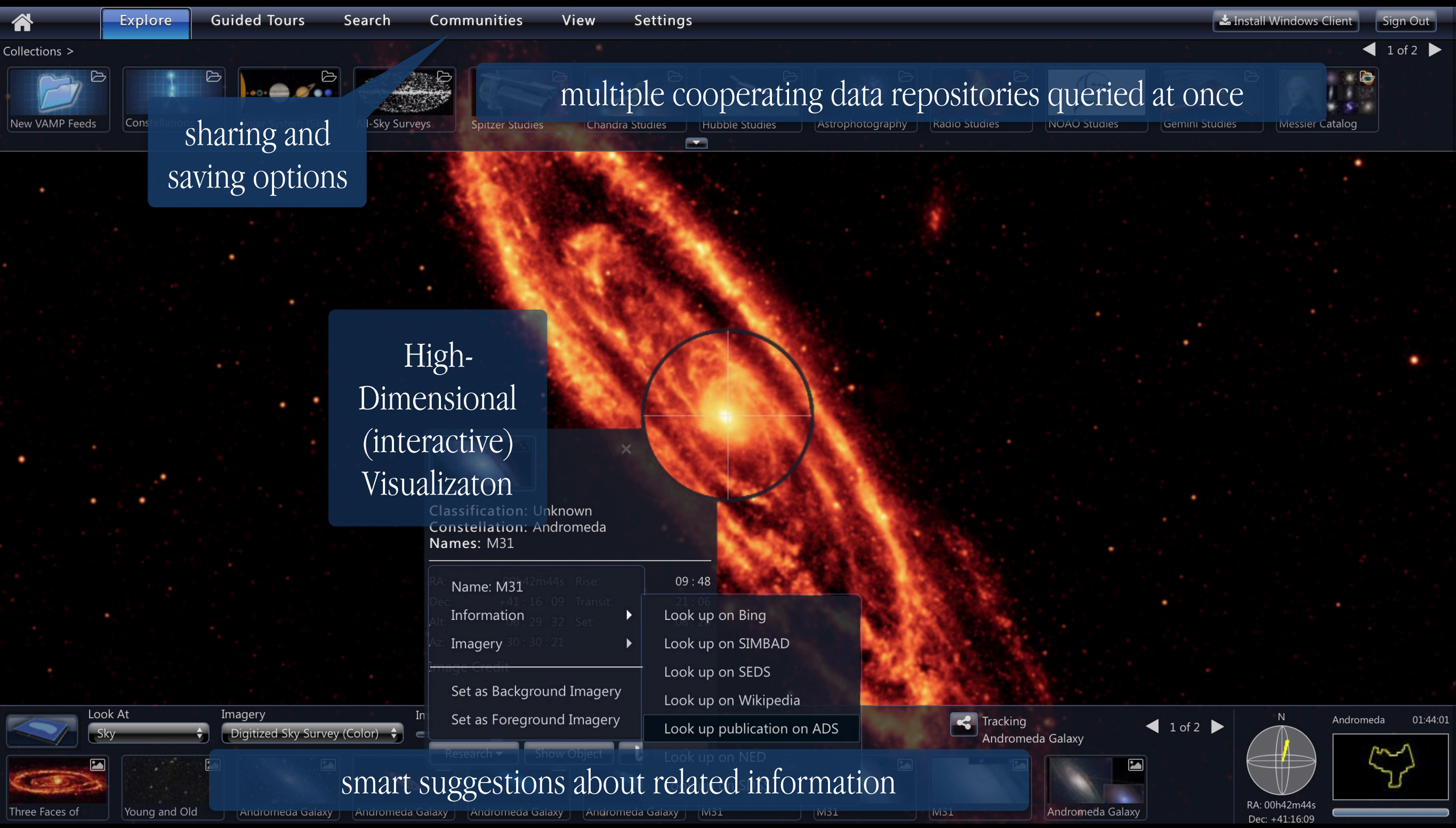
CDS
Centre de Données astronomiques de Strasbourg

SIMBAD

https://www.cfa.harvard.edu/~agoodman/seamless/

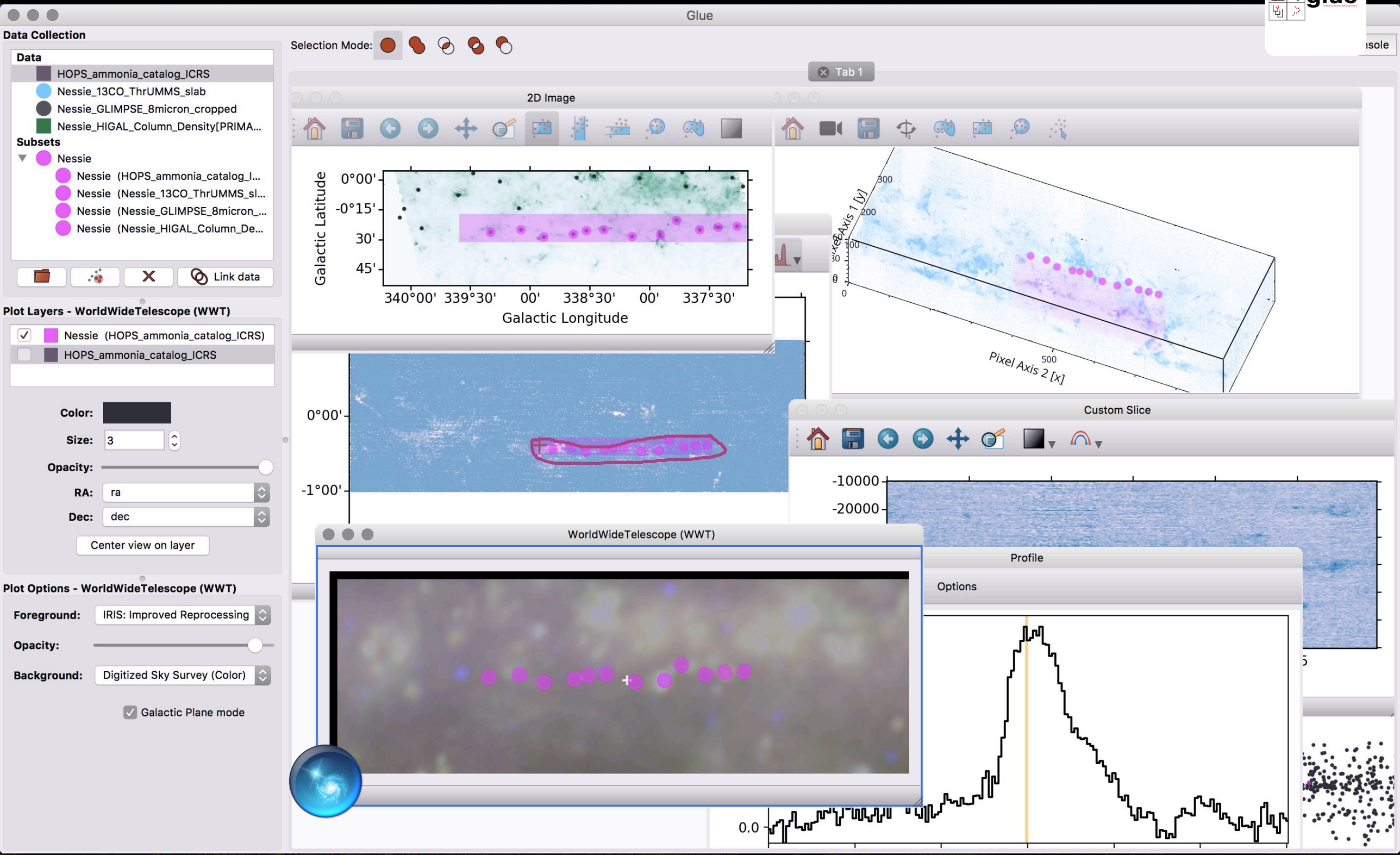
Supported by Microsoft Research NSF NASA Smithsonian

How to make WWT more of a research tool?



http://worldwidetelescope.org/webclient/?loggedIn=true#/place=0.389&cf=77&ra=0.71231&dec=41.26917&fov=1.73327

Dip it in glue.



glueing datasets

glueing data displays

glueing tools

your handout

glue

multidimensional data exploration

built-in

standard data loaders

glueing datasets

Data File 1

Data File 2

define new variables, import/export insights, interactive plots for the web, save state, all from GUI

glueing data displays

standard 1D, 2D & 3D plots

plug-in

custom data loaders

Data File N

link data files' attributes

highlight live or algorithmic selections with Boolean logic

glueing tools

custom buttons, features

custom plots

?

+options

user config.py file (loaders, colors, plot types, +)

IP[y]:

access to all matplotlib functions through built-in IPython terminal

jupyter

run & interact with glue from Jupyter notebook & other tools

glueviz.org

glueing
data displays

glueing
tools

glueing
datasets

*credit for the term “wide data” to Dr. Chris Beaumont original glue developer, co-creator

glueing

data displays

Linked Views

Tukey, 1970s

glueing

tools

glueing

datasets

*credit for the term “wide data” to Dr. Chris Beaumont original glue developer, co-creator

glueing

data displays

Linked Views

Tukey, 1970s

glueing

tools

Open-source, Mashups

~2000s

glueing

datasets

*credit for the term “wide data” to Dr. Chris Beaumont original glue developer, co-creator

glueing

data displays

Linked Views

Tukey, 1970s

glueing

tools

Open-source, Mashups

~2000s

glueing

datasets

Wide Data*

e.g. glue, now

*credit for the term “wide data” to Dr. Chris Beaumont original glue developer, co-creator

glueing

data displays

Linked Views

Tukey, 1970s

glueing

tools

Open-source, Mashups

~2000s

glueing

datasets

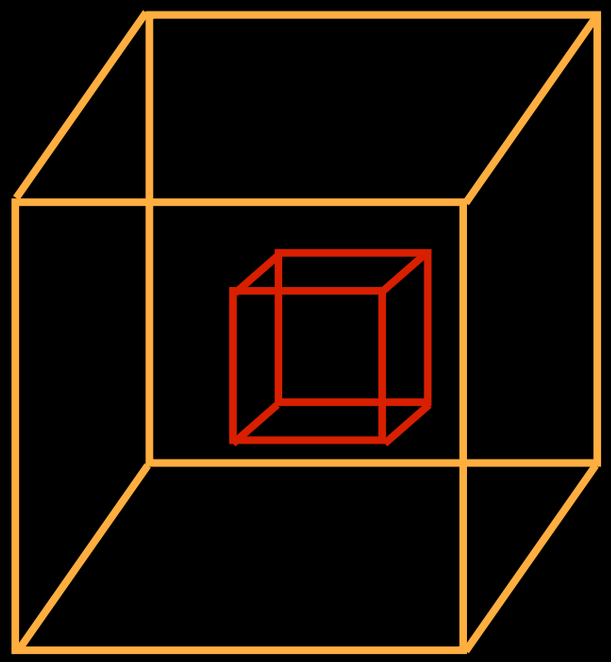
Wide Data*

e.g. glue, now

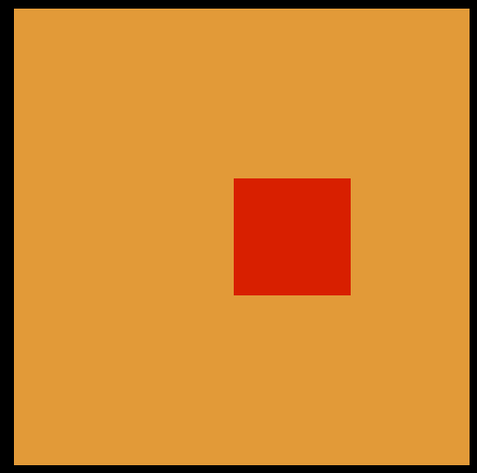
*credit for the term “wide data” to Dr. Chris Beaumont original glue developer, co-creator

Linked Views of High-dimensional Data

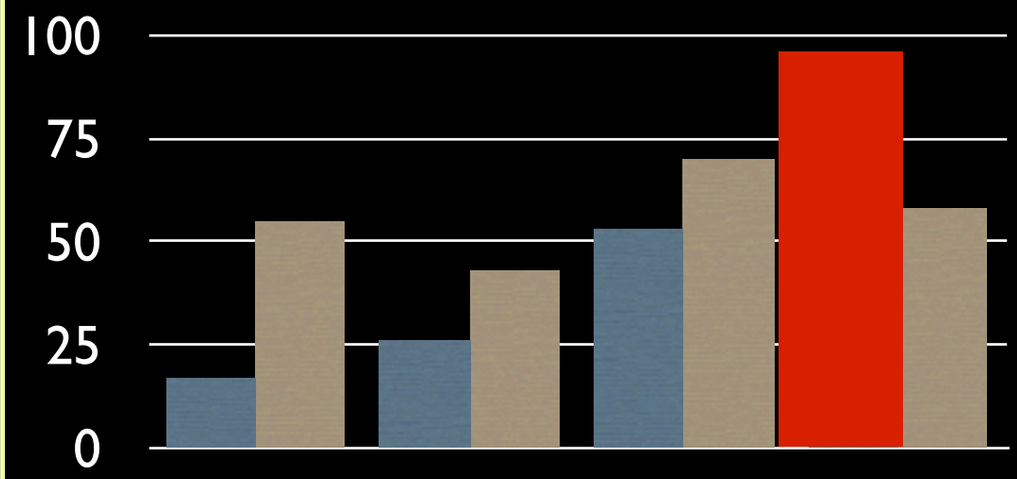
John Tukey



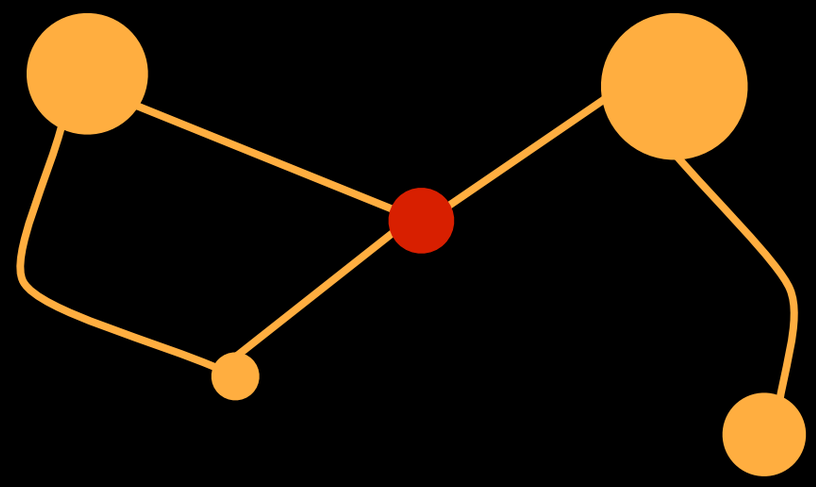
3D



2D

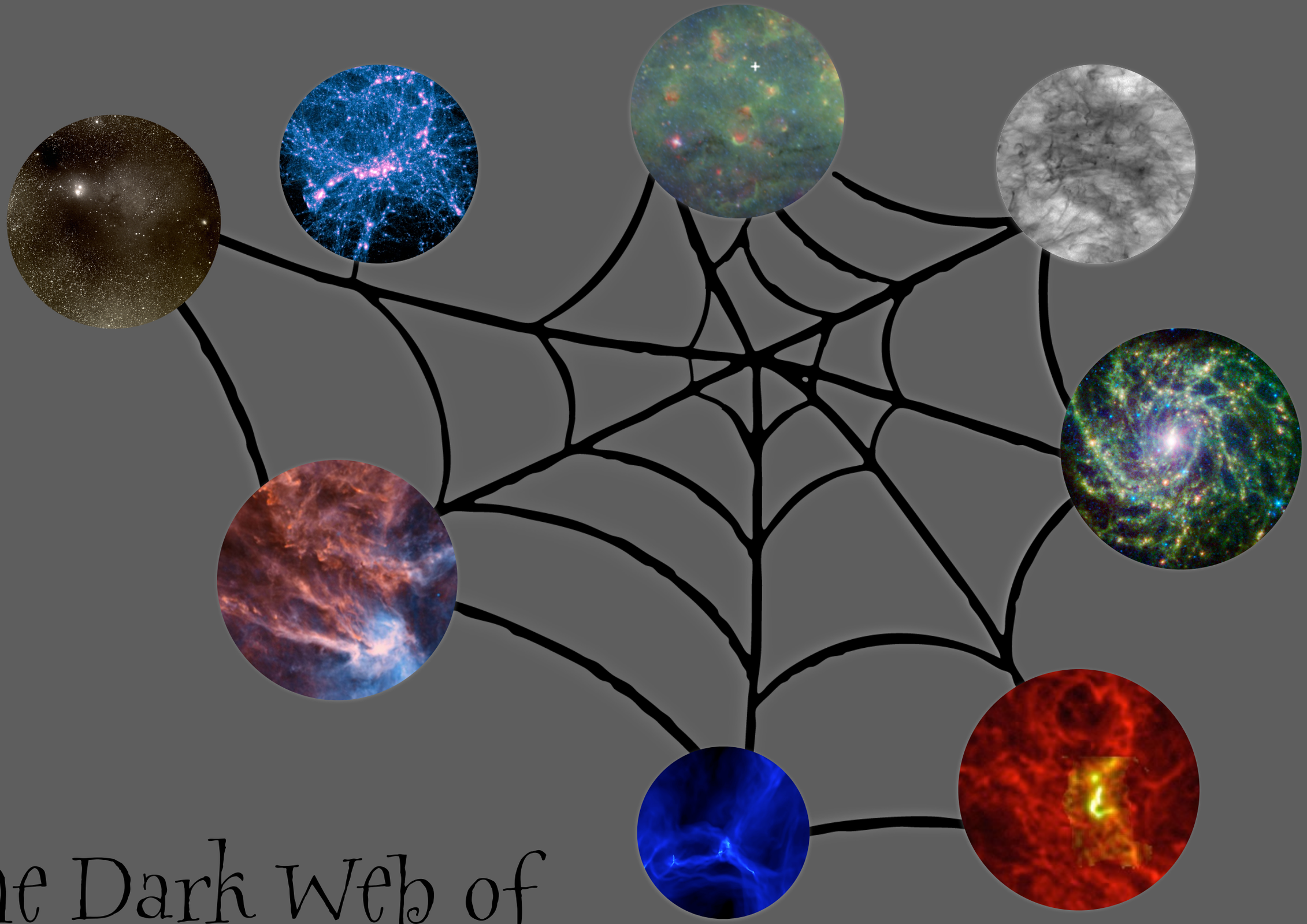


Data Abstraction



figure, by M. Borkin, reproduced from Goodman 2012, "Principles of High-Dimensional Data Visualization in Astronomy"

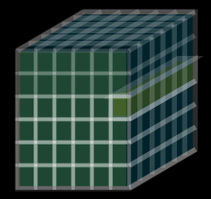
Why high-dimensional data matters to me.

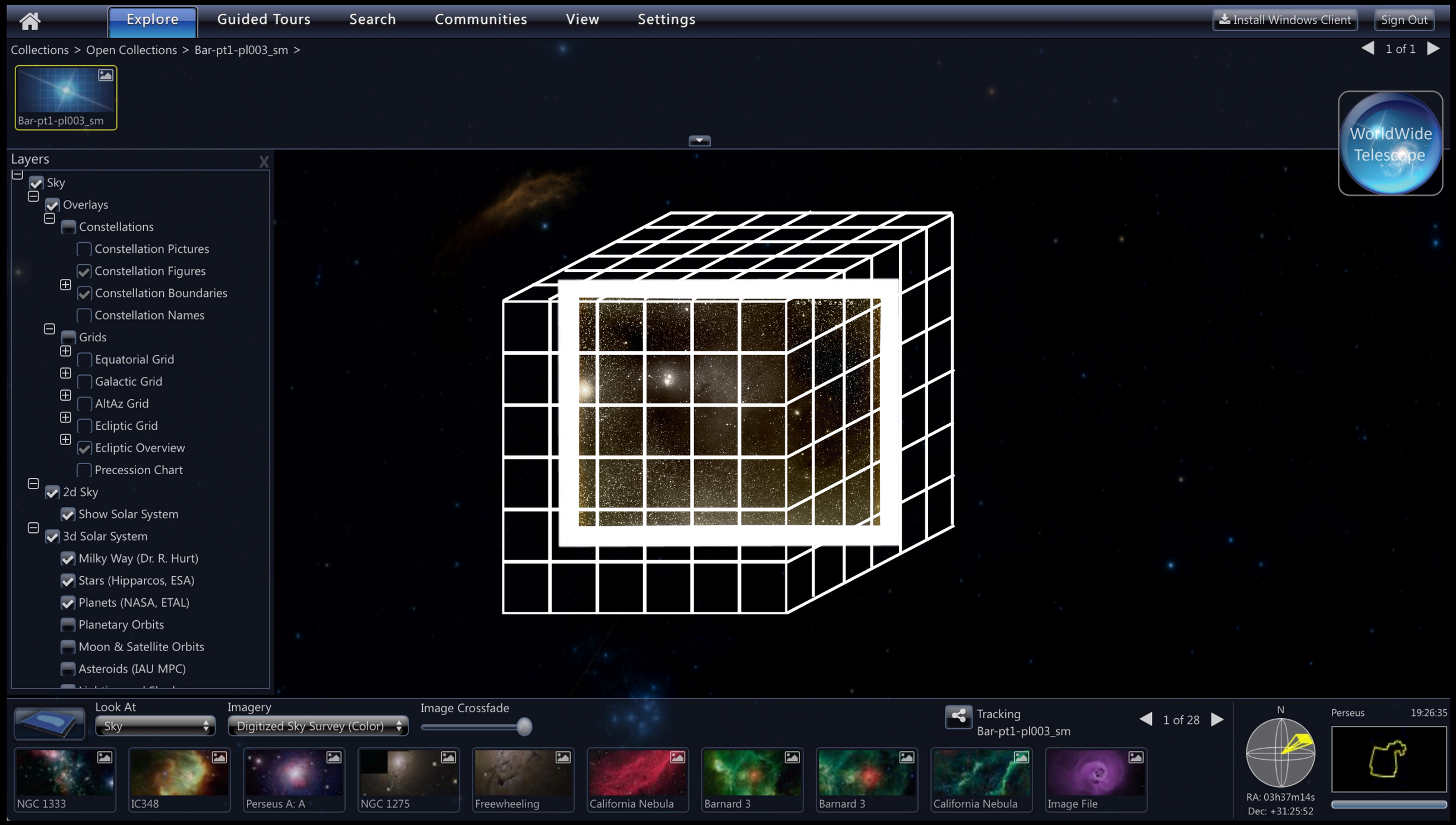


The Dark Web of Star Formation

Alyssa A .Goodman
Harvard-Smithsonian Center for Astrophysics
& Radcliffe Institute for Advanced Study

“Data, Dimensions, Display”





Data, Dimensions, Display

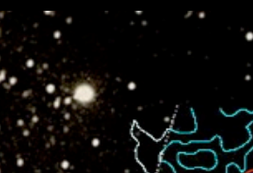
mm peak (Enoch et al. 2006)

sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)

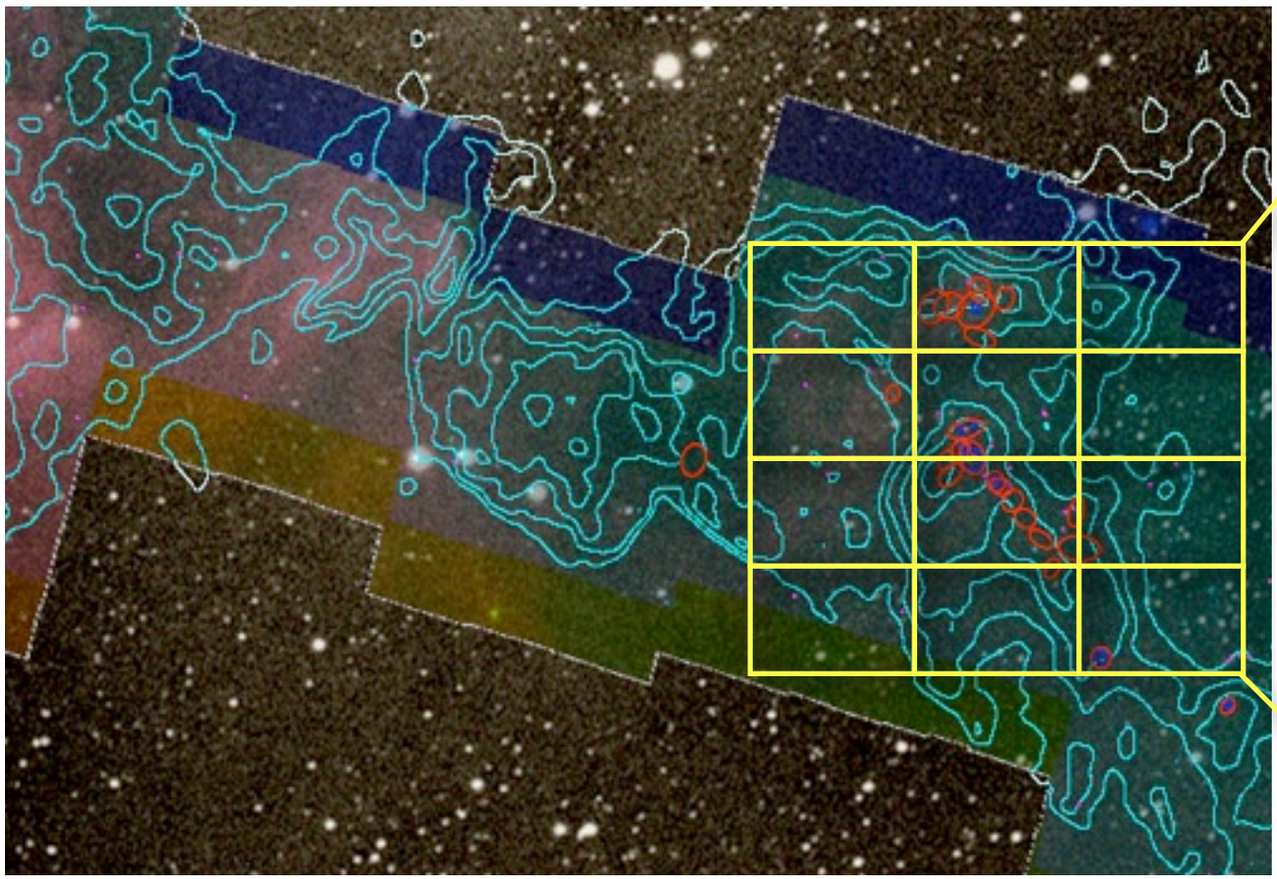
^{13}CO (Ridge et al. 2006)

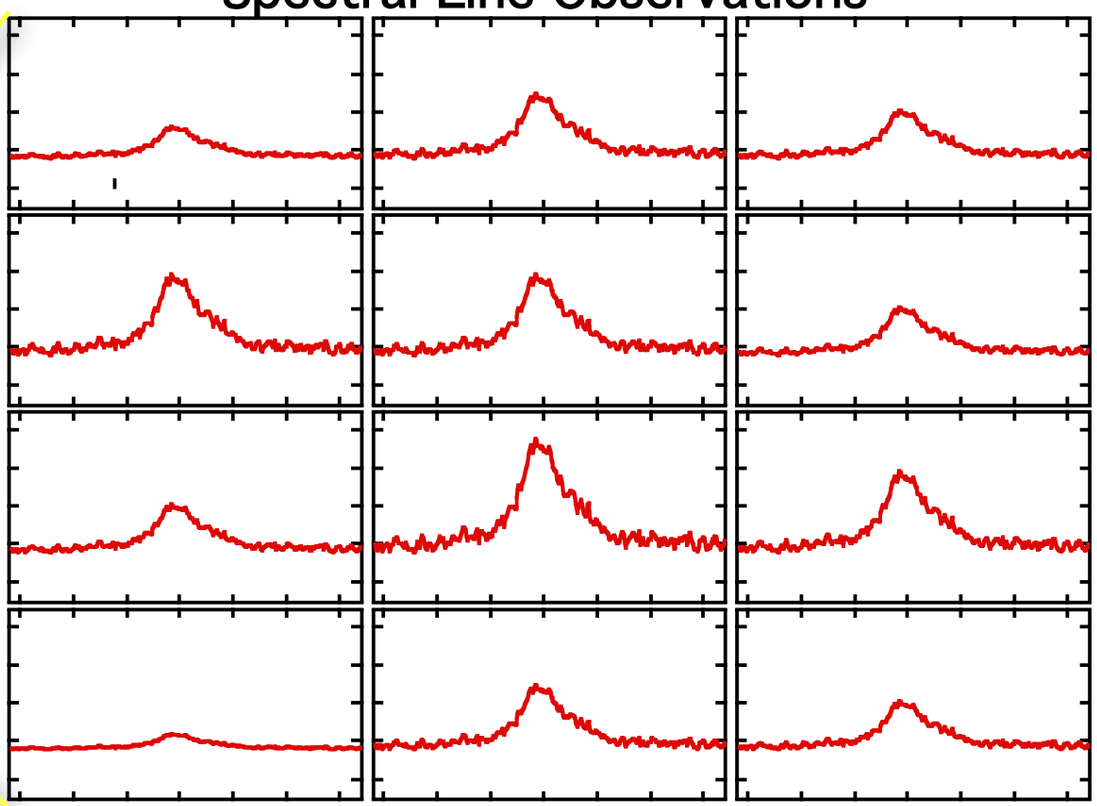
mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al.)

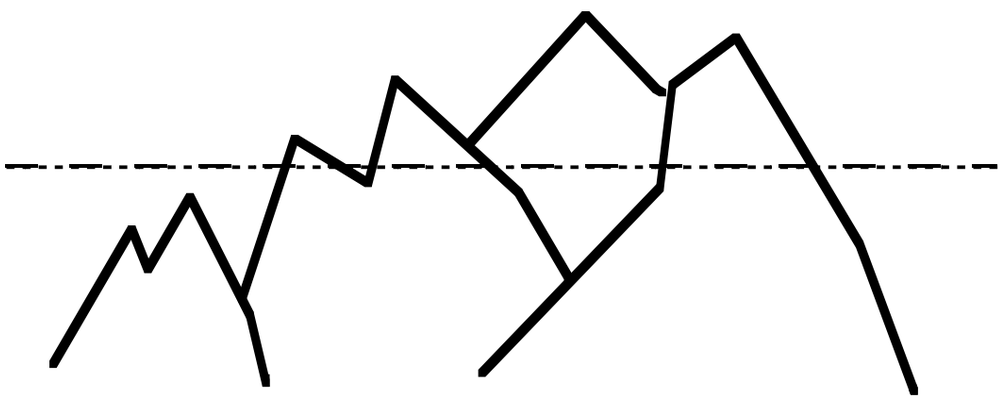
Optical image (Barnard 1927)



Data, Dimensions, Display







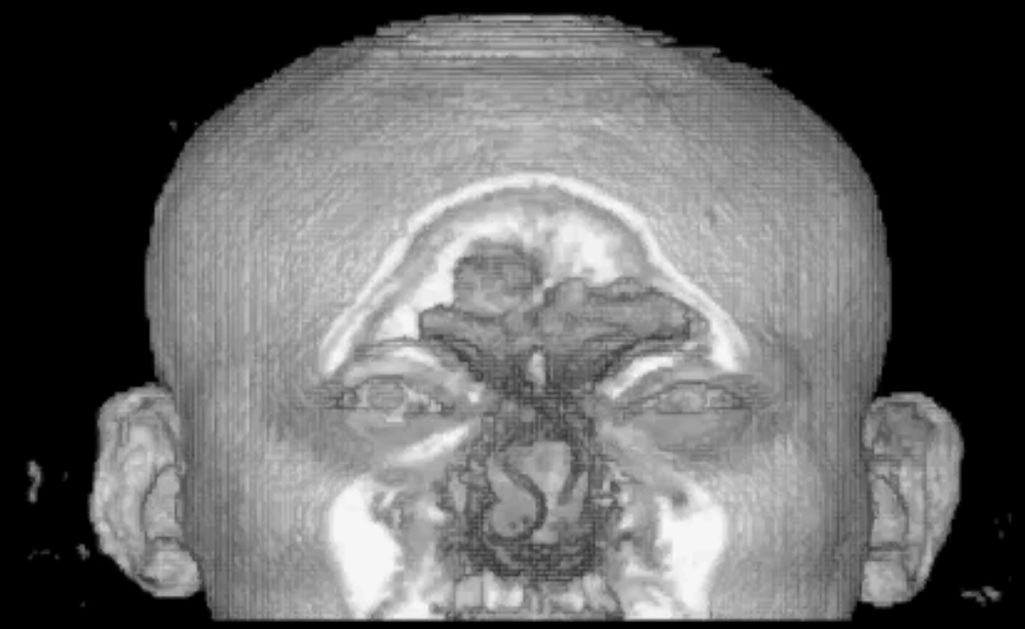
Mountain Range

No loss of information

Loss of 1 dimension

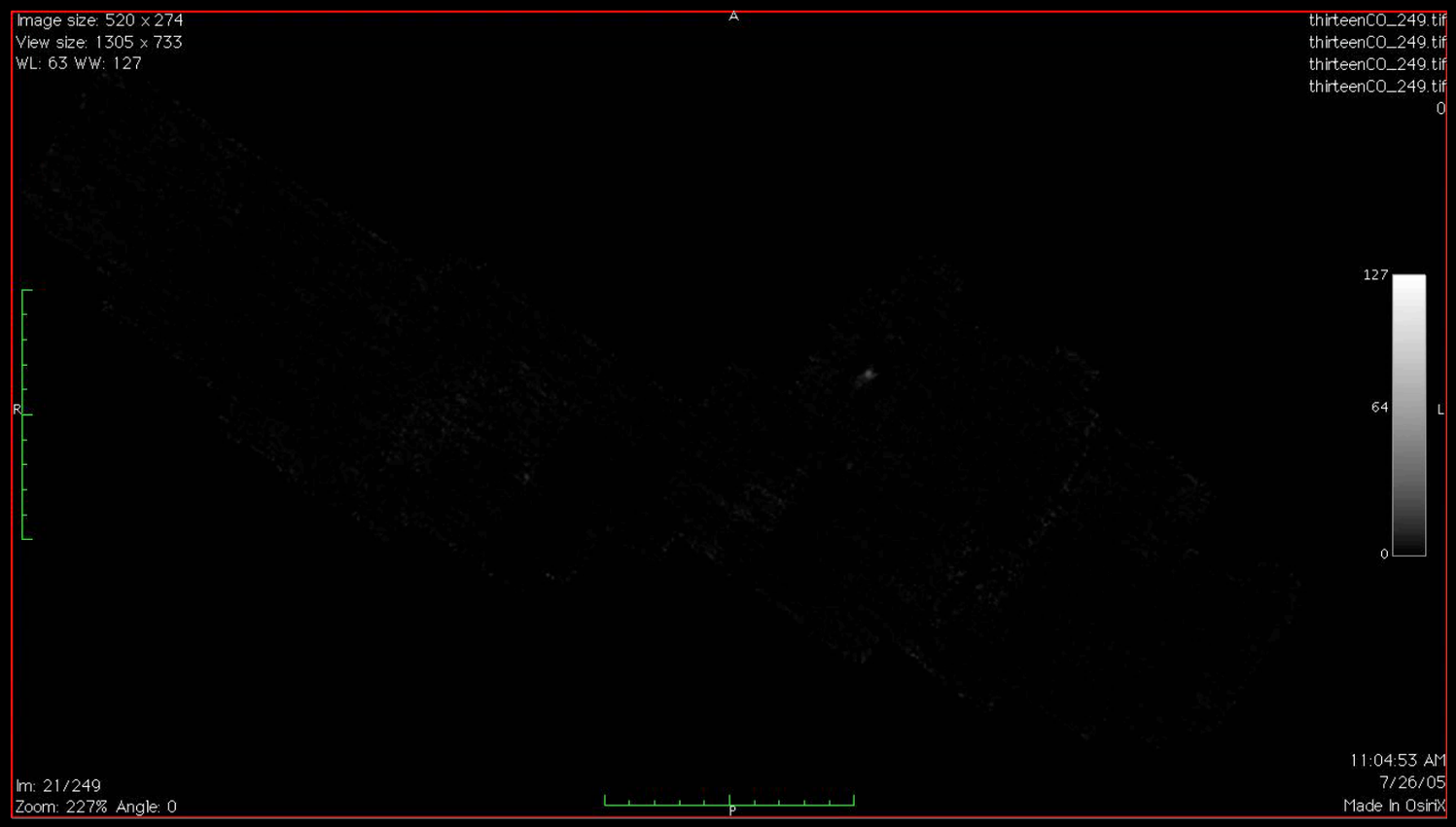
“AstronomicalMedicine”

“KEITH”



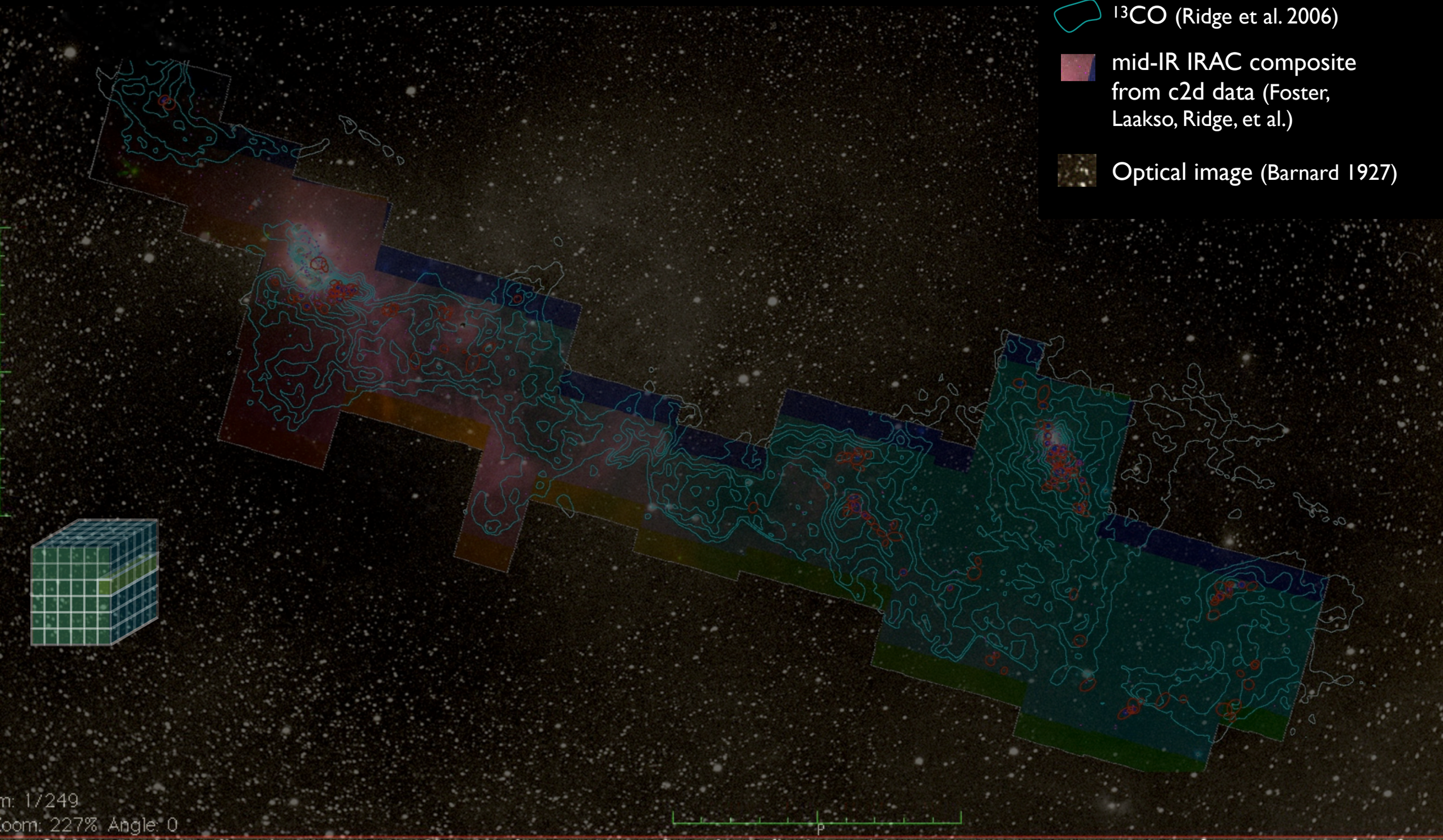
“z” is depth into head

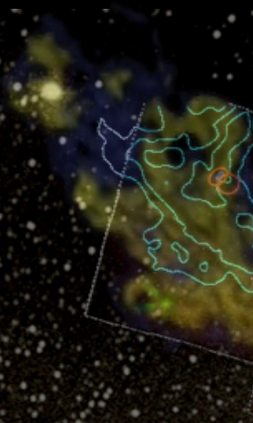
“PERSEUS”



“z” is line-of-sight velocity

Data, Dimensions, Display

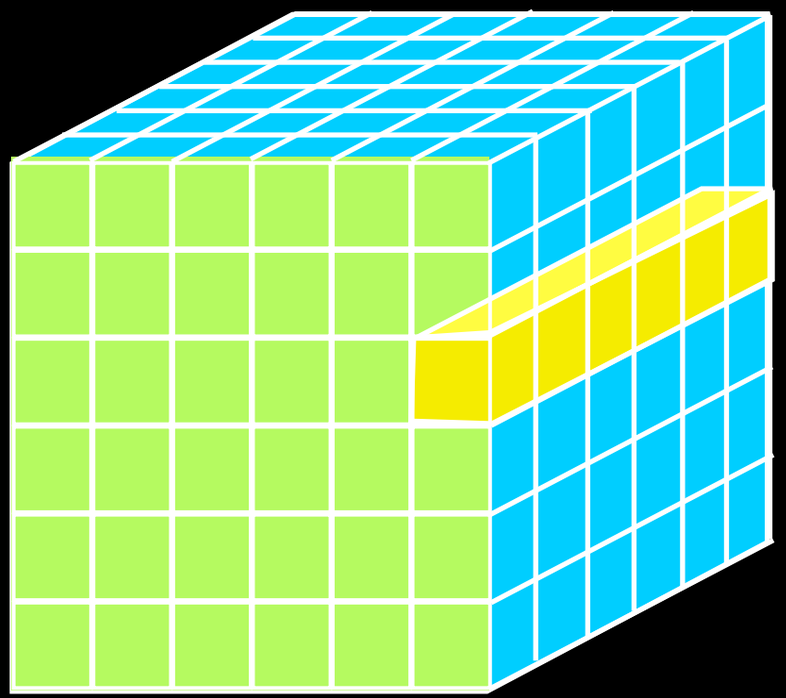




3D Viz made with VolView

AstronomicalMedicine@IIG

COMPLETE

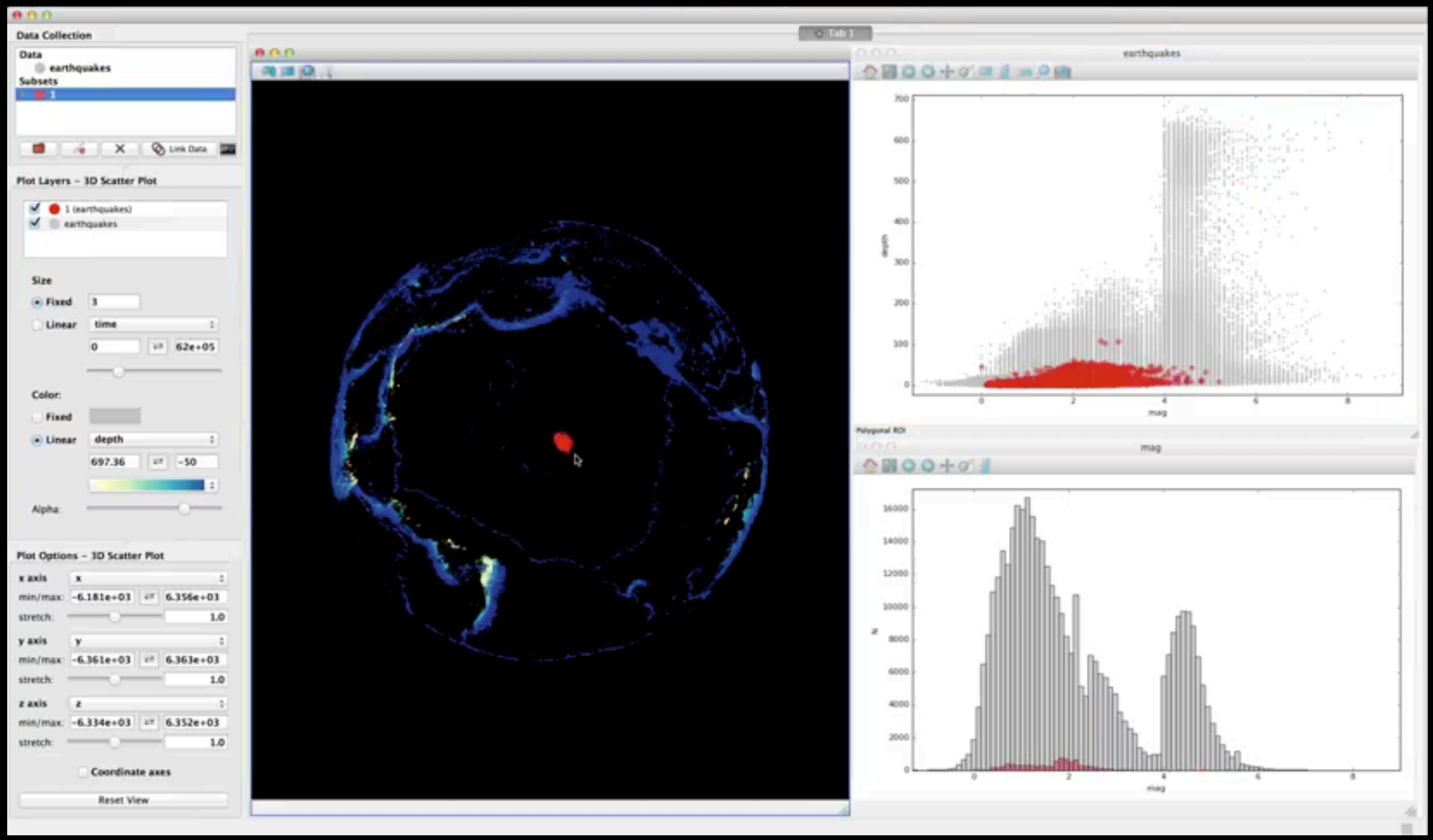


Data, Dimensions, Display

1D: Columns = “Spectra”, “SEDs” or “Time Series”
2D: Faces or Slices = “Images”
3D: Volumes = “3D Renderings”, “2D Movies”
4D: Time Series of Volumes = “3D Movies”

Linked Views of High-dimensional Data (in Python)

glue

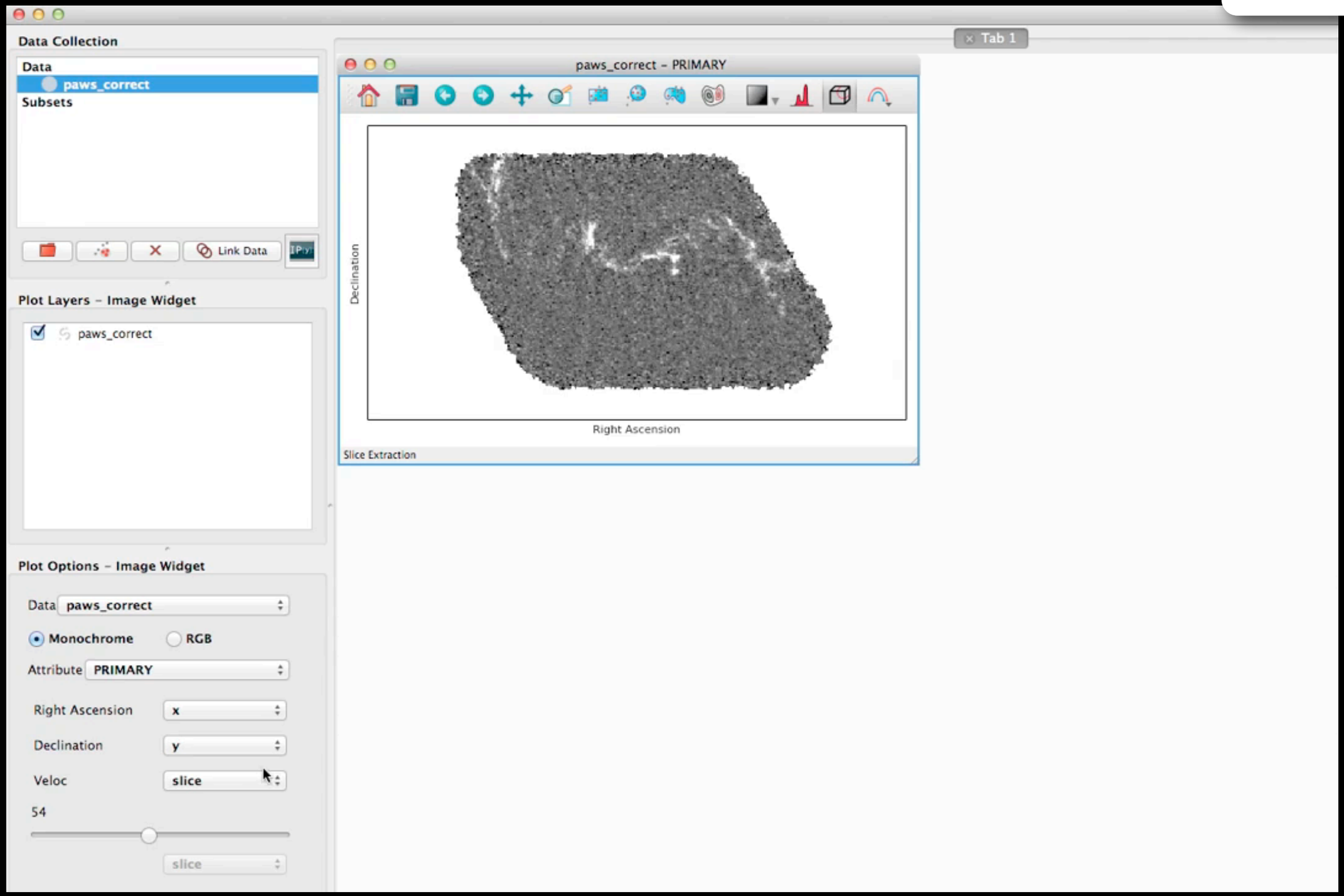


video by Tom Robitaille, lead glue developer

glue created by: C. Beaumont, M. Borkin, P. Qian, T. Robitaille, and A. Goodman, PI

Linked Views of High-dimensional Data (in Python)

glue



video by Chris Beaumont, glue developer

glue created by: C. Beaumont, M. Borkin, P. Qian, T. Robitaille, and A. Goodman, PI

in

Glue

Data Collection

Selection Mode:

View Console

−50.0 −46.7 −43.3 −40.0 −36.7 −33.3 −30.0

V_{LSR} (km s^{-1})

[Z_0=25.0 pc, R_0=8.5 kpc, Θ_0=220 km/s]

no tilt of plane

for more on the "Skeleton" of the Milky Way
see Goodman et al. 2014, Zucker et al. 2015, 2017

Galactic Latitude (b)

1.0 0.5 0.0 −0.5 −1.0

340 338 336 334

Galactic Longitude (l)

in glue

Glue

Data Collection

Data
- Nessie_HIGAL_Column_Density[PRIMARY]
- HOPS_dense_gas_catalog

Subsets
- Nessie_Boundary

Link data

Plot Layers

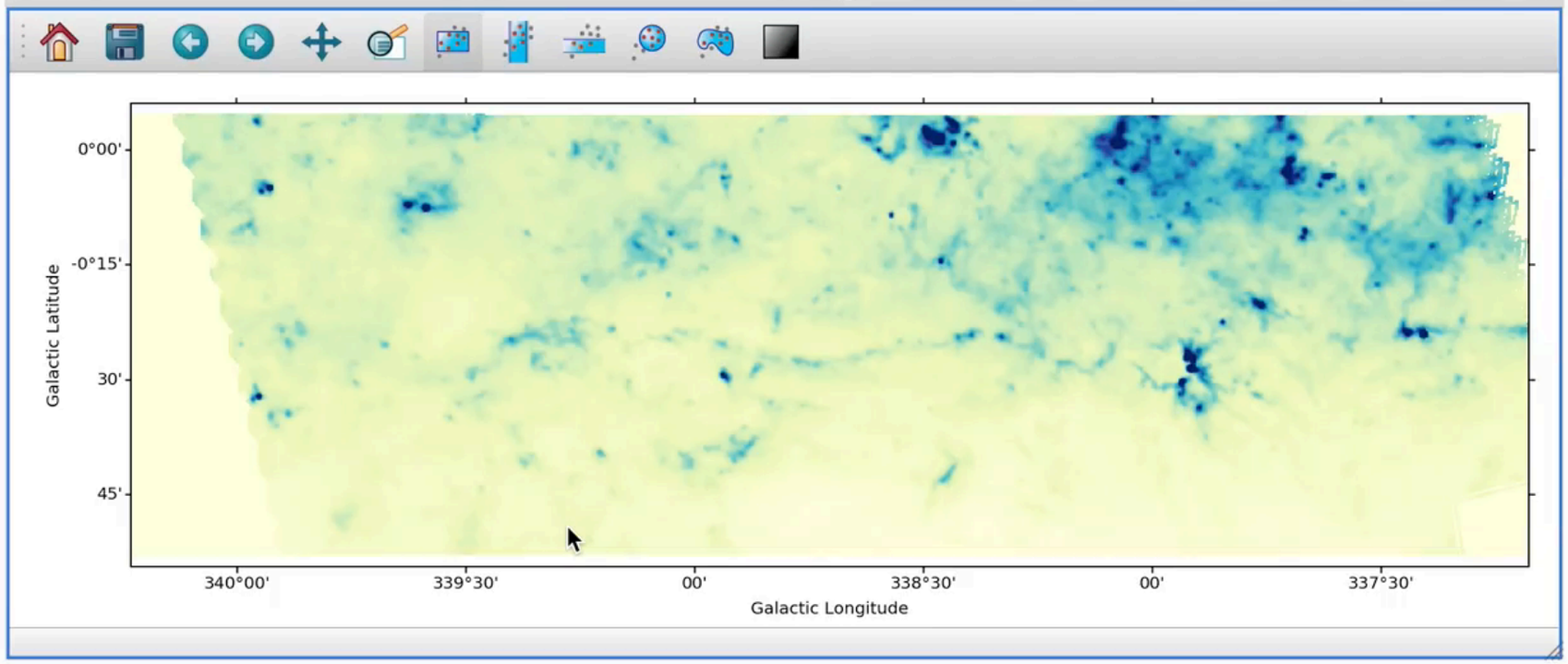
Plot Options

Selection Mode:

View Console

Tab 1

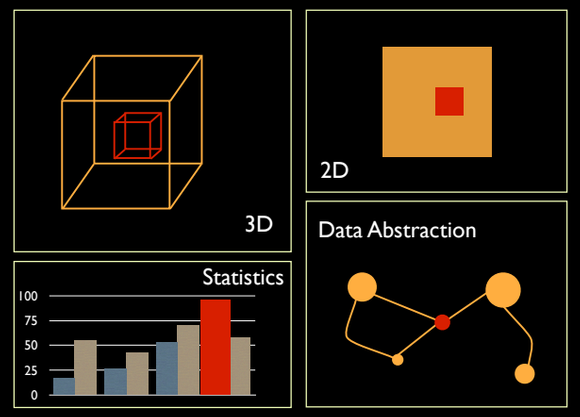
2D Image



glueing
data displays

Linked Views

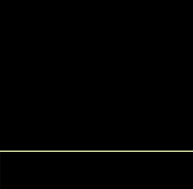
Tukey, 1970s



glueing
tools

Open-source, Mashups

~2000s

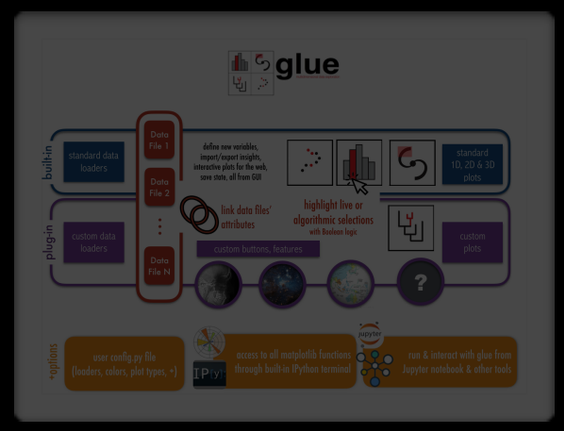


in

glueing
datasets

Wide Data*

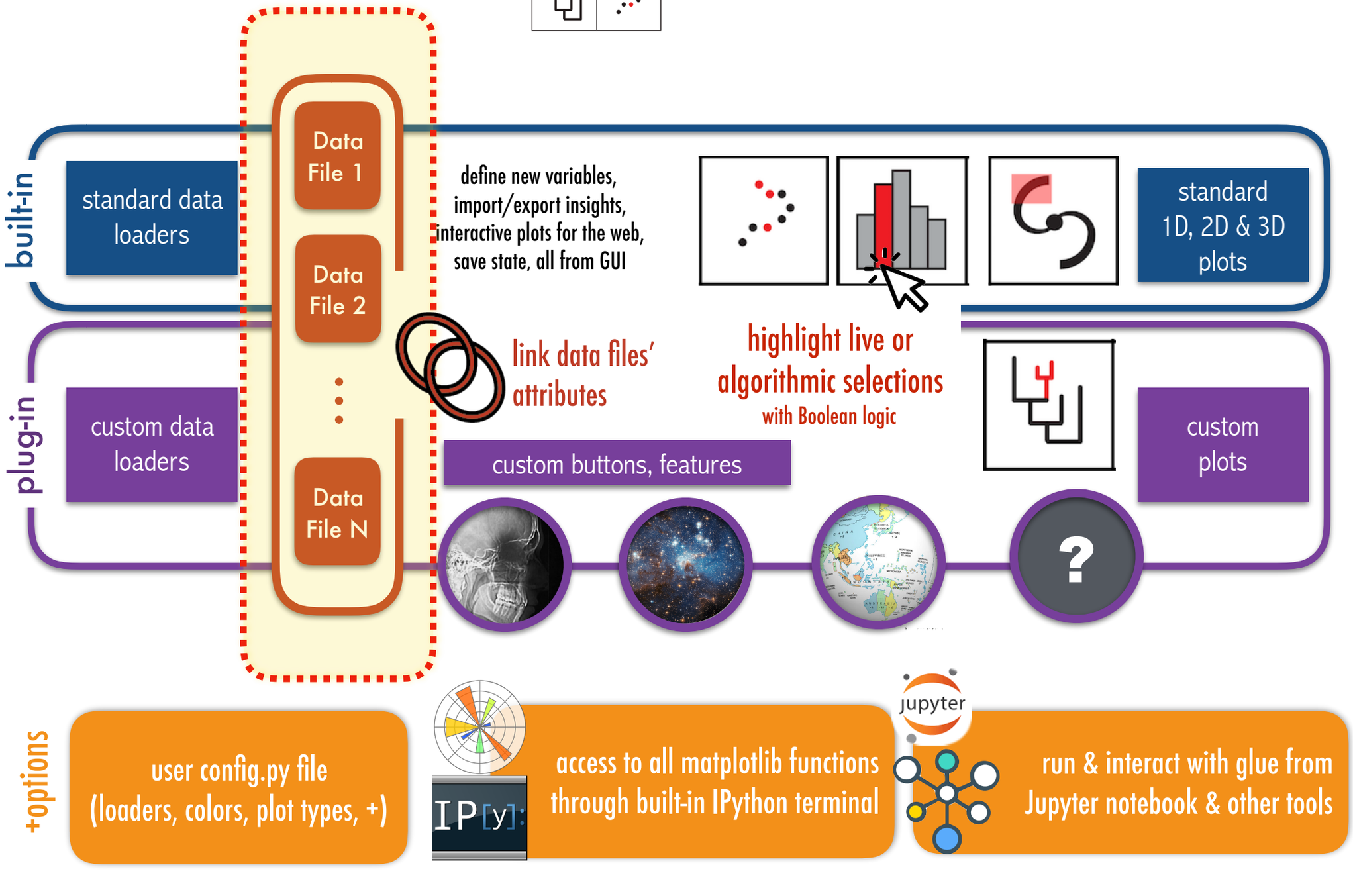
e.g. glue, now



*credit for the term “wide data” to Dr. Chris Beaumont original glue developer, co-creator

glue

multidimensional data exploration



glueviz.org

Wide Data

COMPLETE

mm peak (Enoch et al. 2006)

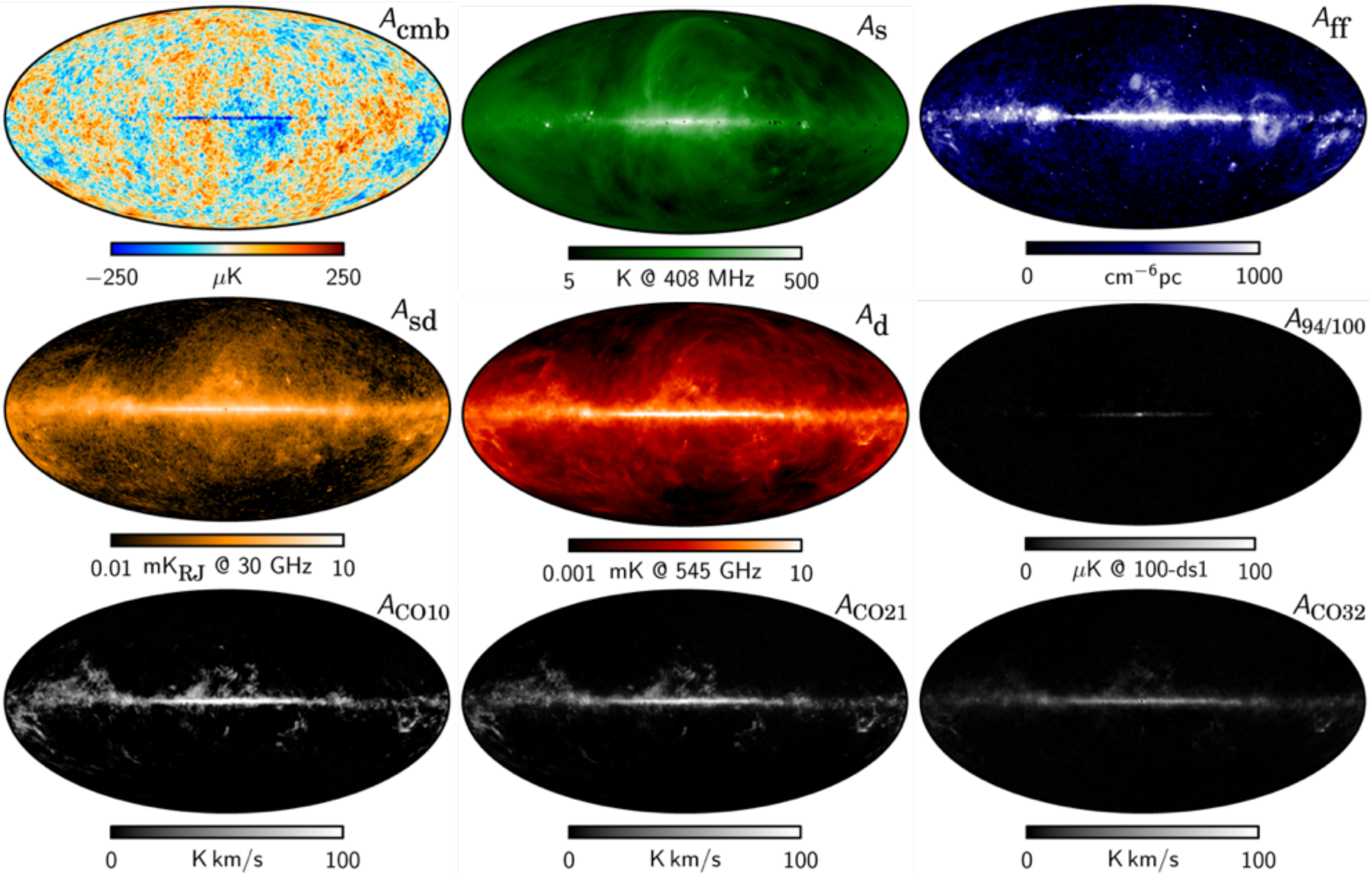
sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)

^{13}CO (Ridge et al. 2006)

mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al.)

Optical image (Barnard 1927)

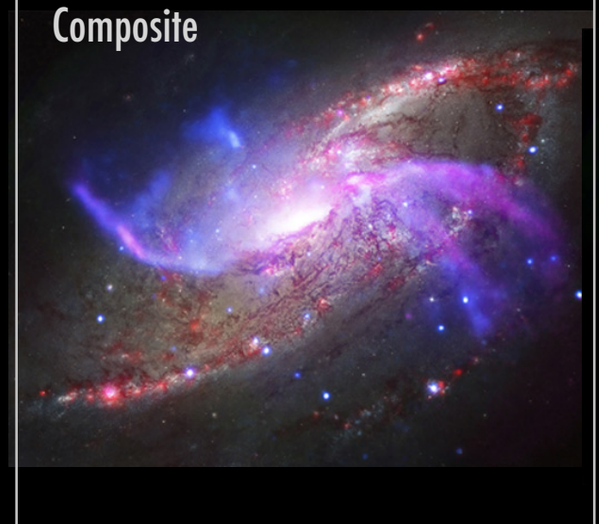
Wide Data

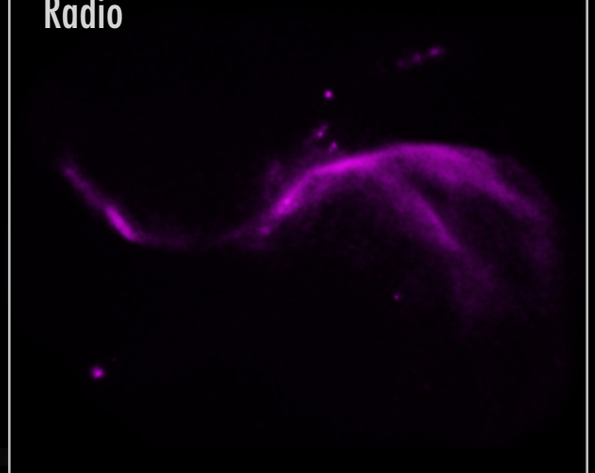


Temperature Foreground amplitudes from Commander, Planck Data [Feb 2015]

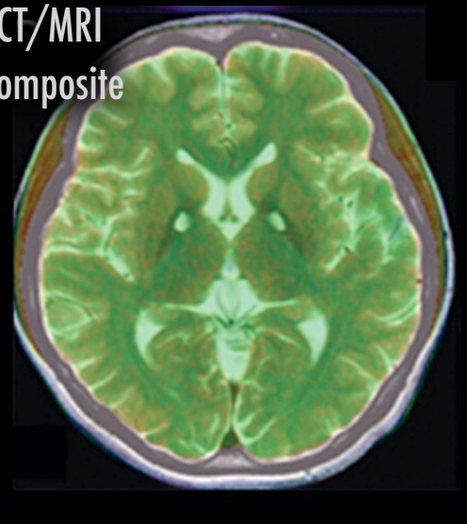
Wide Data

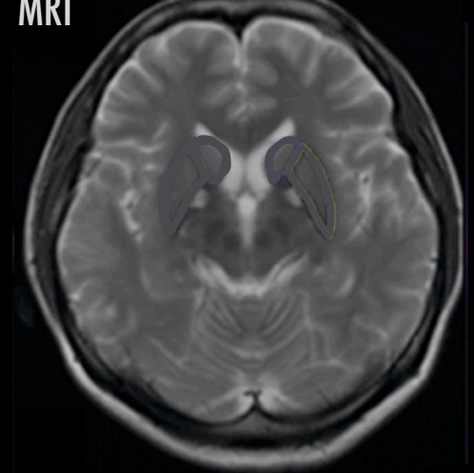
AstronomicalMedicine

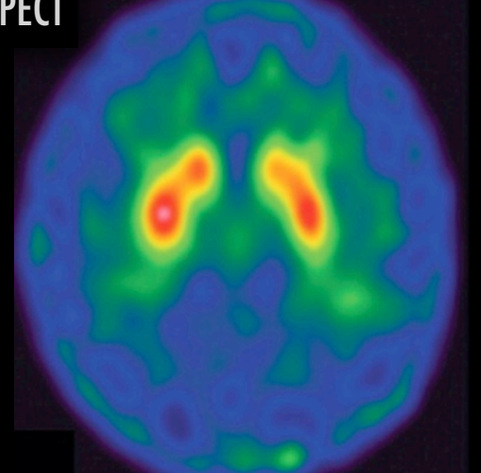




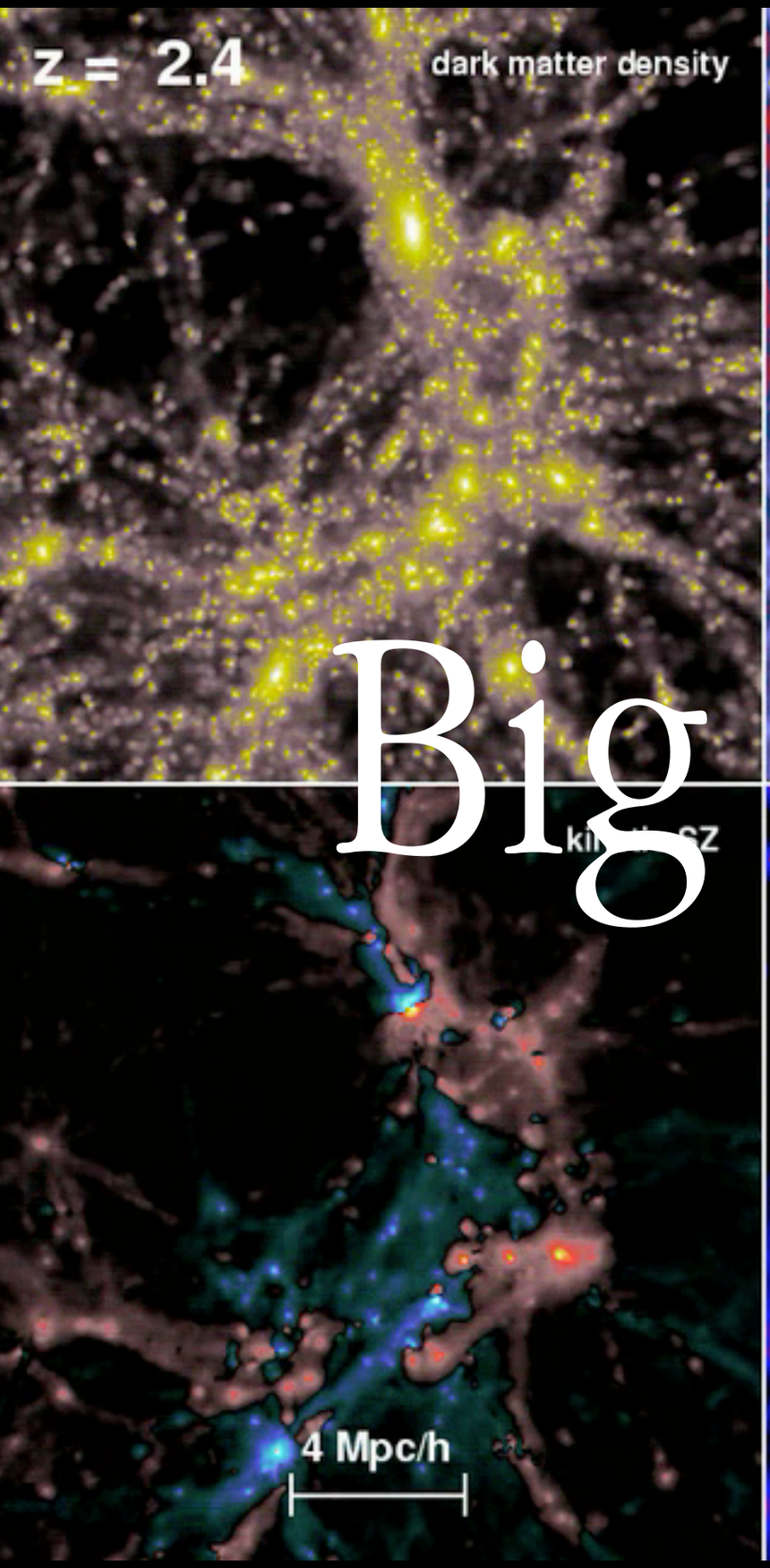
chandra.harvard.edu/photo/2014/m106/

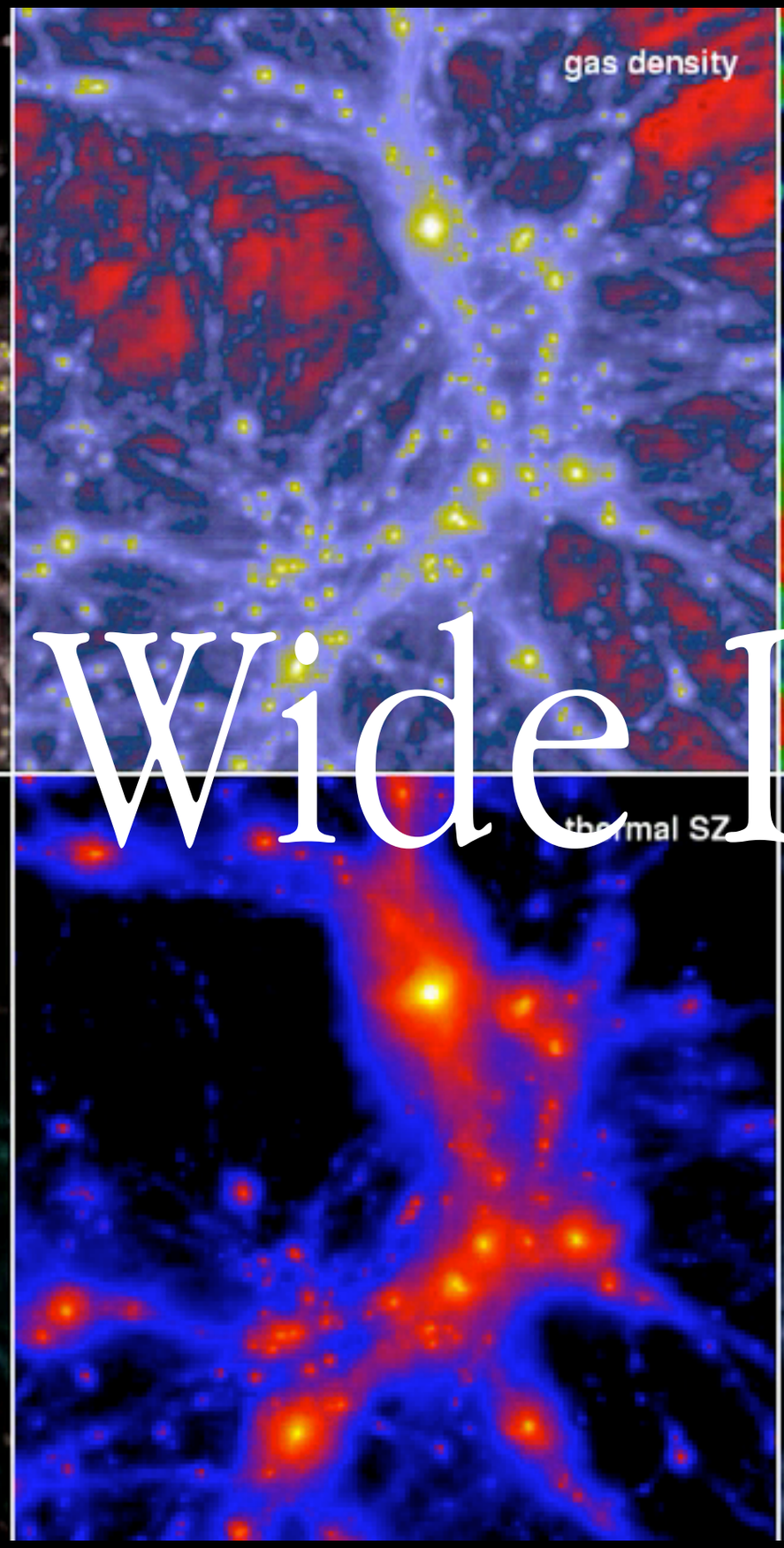


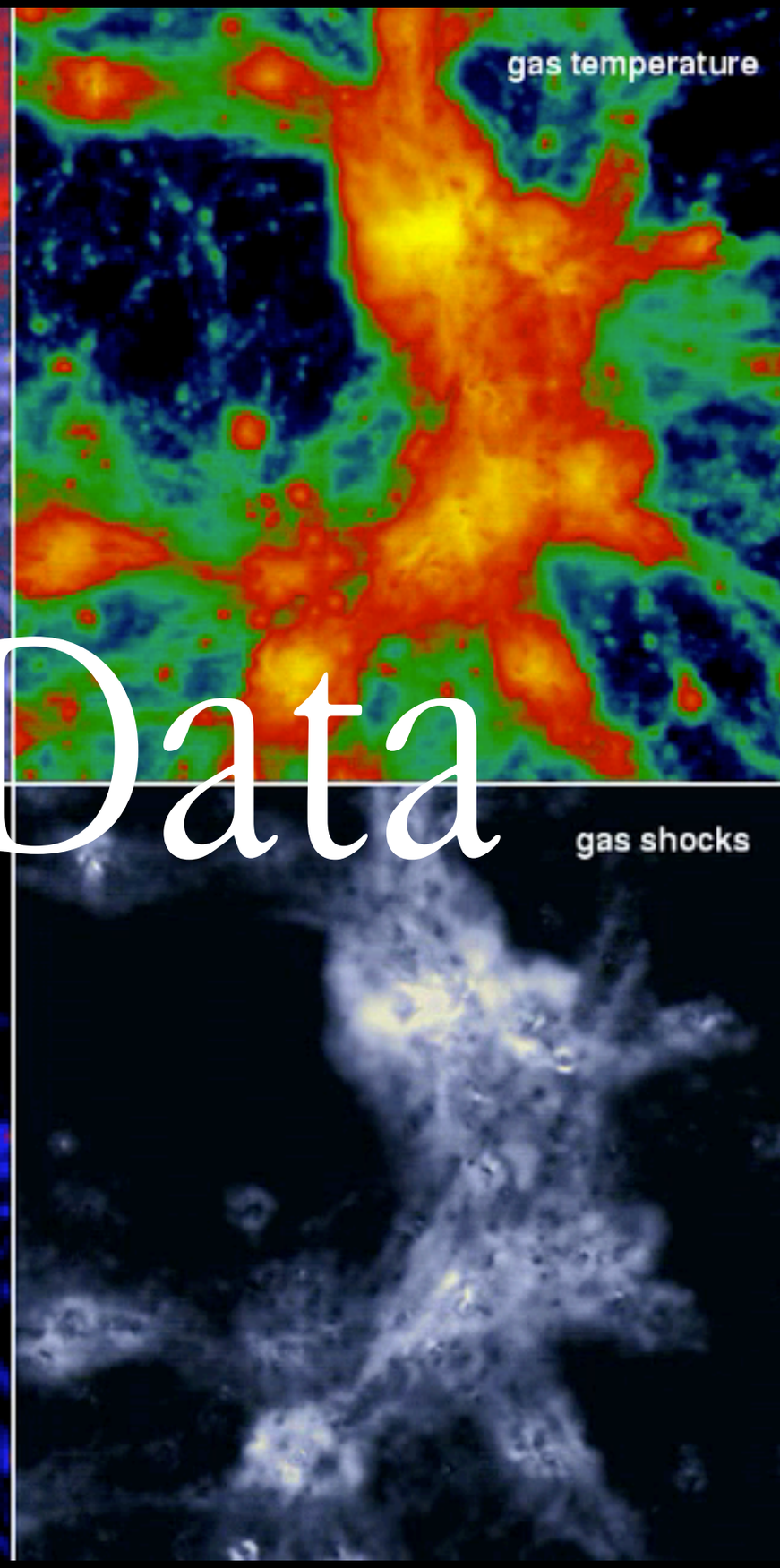




Chang, et al. 2011, brain.oxfordjournals.org/content/134/12/3632

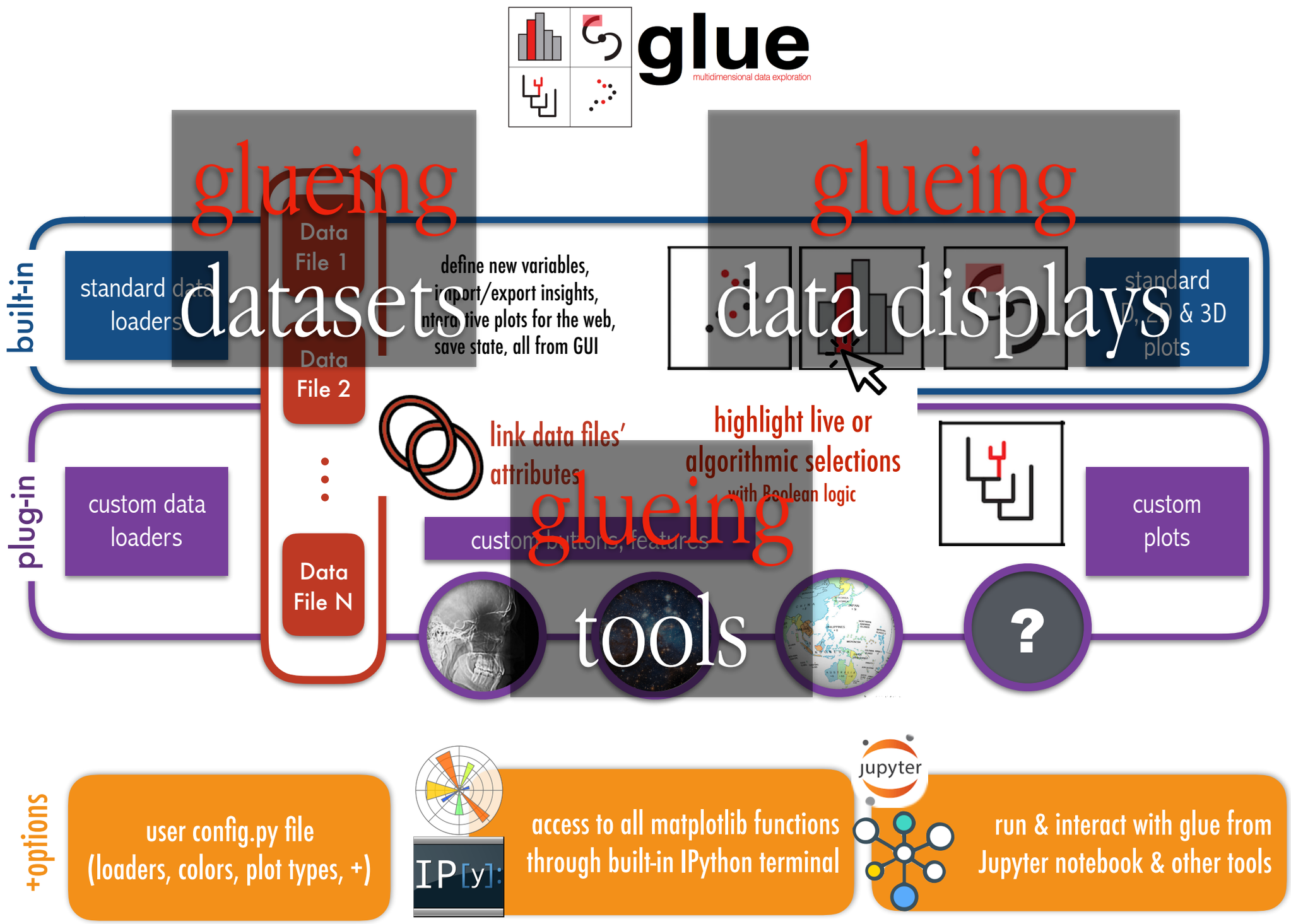






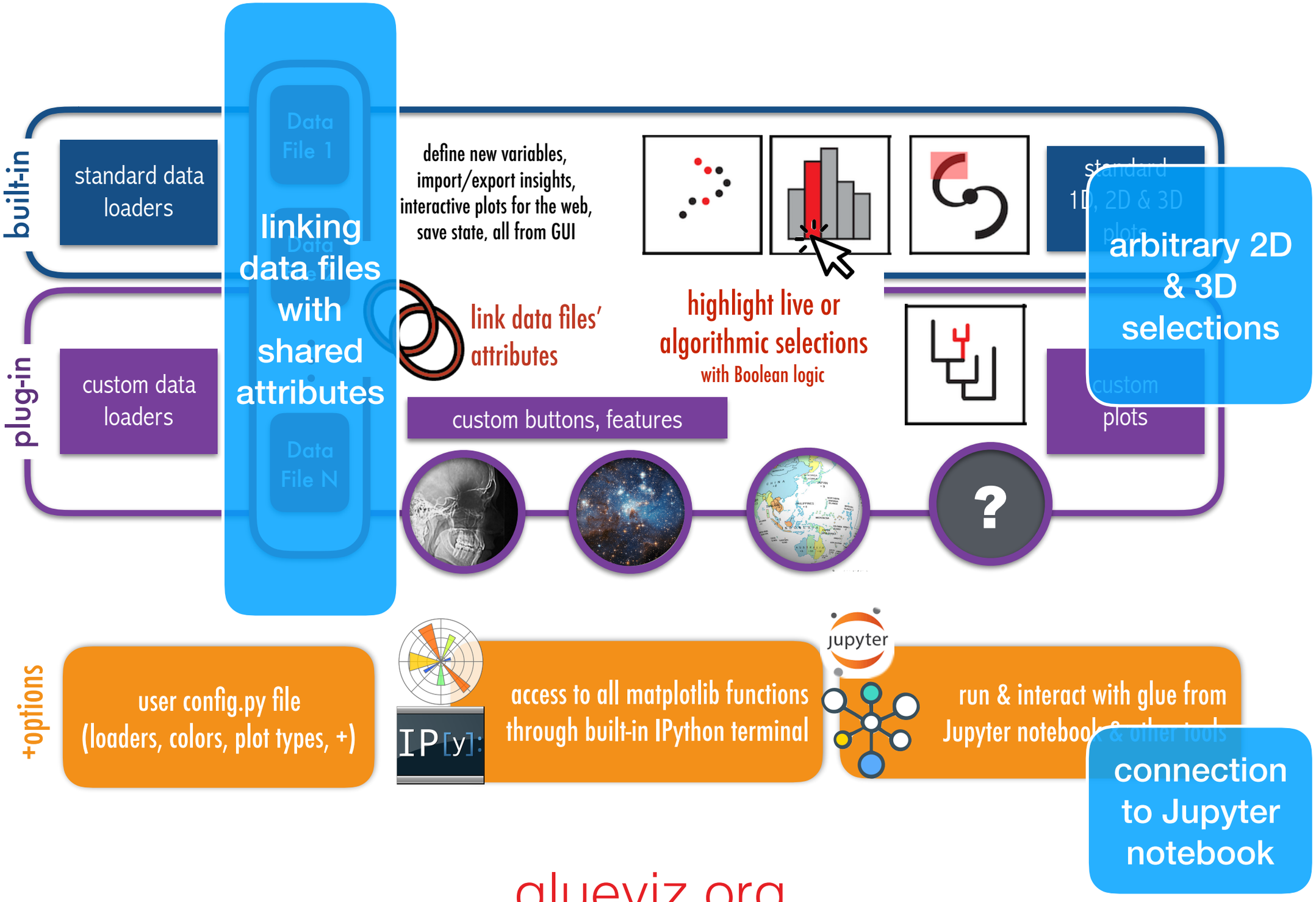
Big Wide Data

Movie: Volker Springel, formation of a cluster of galaxies. Millenium Simulation requires 25TB for output.


glue
multidimensional data exploration
glueing
glueing
datasets
data displays
built-in
standard data loaders
Data File 1
Data File 2
Data File N
define new variables, import/export insights, interactive plots for the web, save state, all from GUI
standard 1D, 2D & 3D plots
plug-in
custom data loaders
link data files' attributes
highlight live or algorithmic selections with Boolean logic
custom buttons, features
glueing
custom plots
tools
?
+options
user config.py file (loaders, colors, plot types, +)
IP[y]:
access to all matplotlib functions through built-in IPython terminal
jupyter
run & interact with glue from Jupyter notebook & other tools

glueviz.org

Which parts of glue are novel?



glueviz.org

What's next . . .

arbitrary 2D & 3D selections

(true) 3D Selection

linking data files with shared attributes

(remote accesss to) Big (and Wide) Data

connection to Jupyter notebook

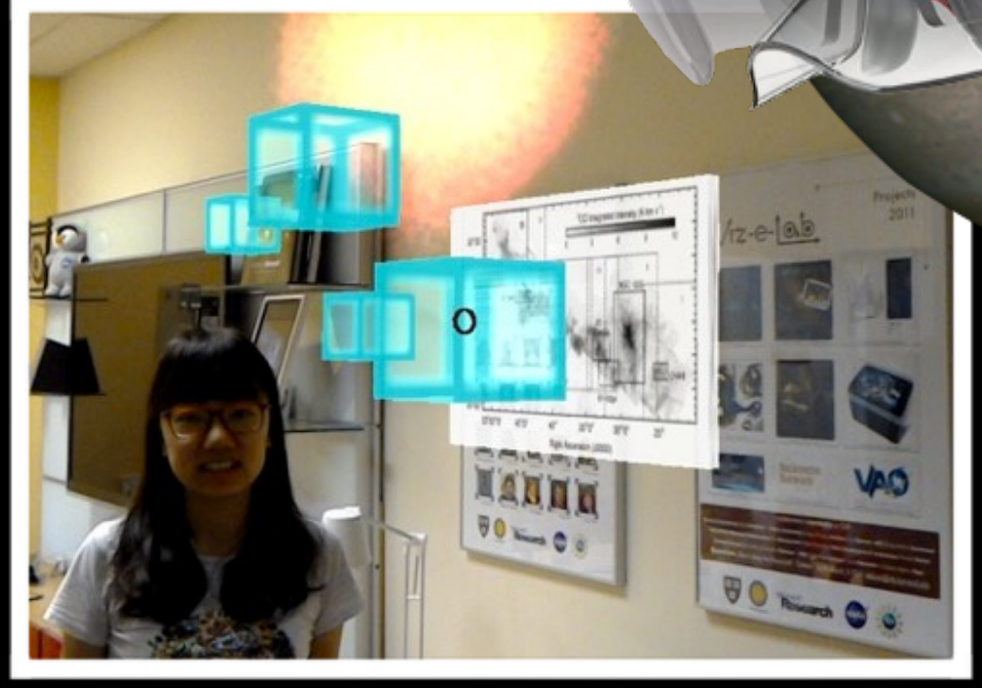
Bridges to Web (e.g. Jupyter Labs, bokeh)

(more) publishing options

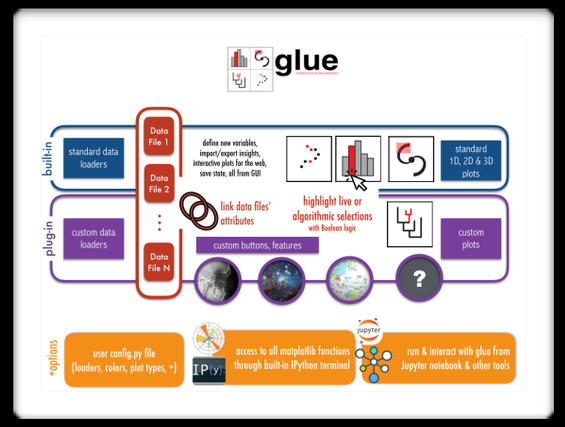
more partners (R, GIS, networks, etc. . .)

The challenge of 3D Selection

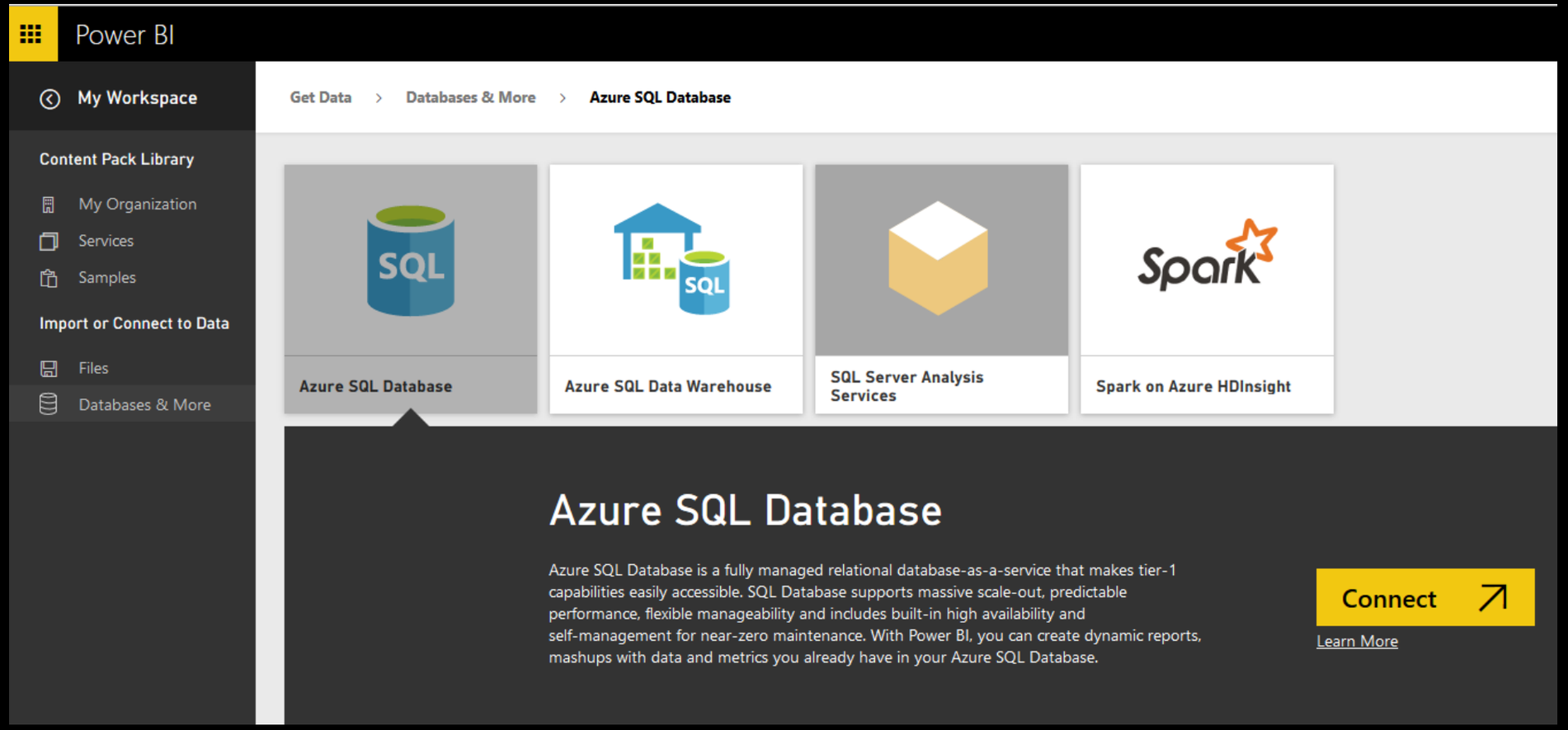
Viz-e-Lab



Remote Access to Big & Wide Data



+data abstraction layer (e.g. via Dataverse)
+replace matplotlib with OpenGL-backed 3D viewer
+data shaders (e.g. bokeh)

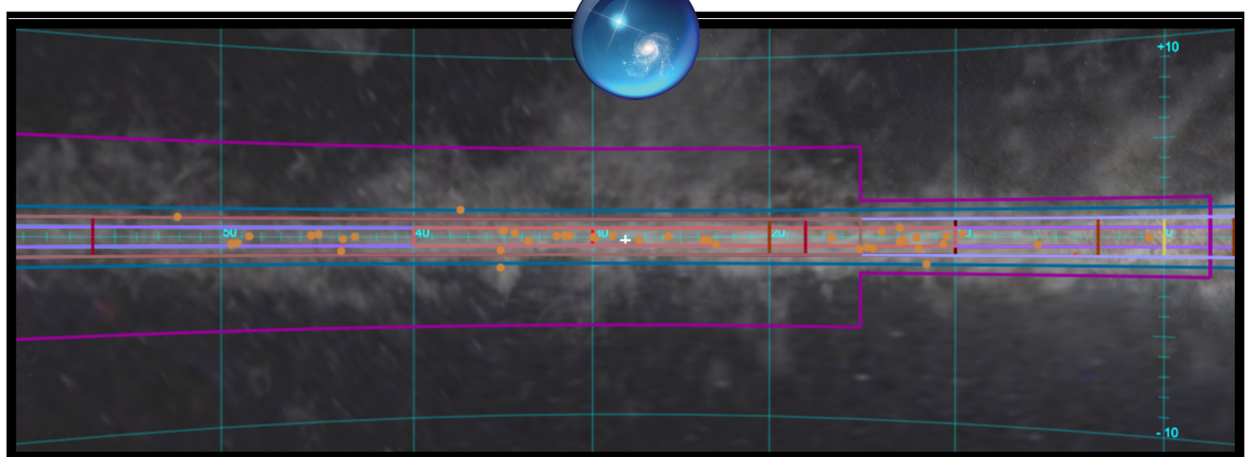


Remote Access & Dataverse, e.g. milkyway3d.org

Milky Way 3D

Galactic Plane Coverage Tool

MilkyWay3D.org is a tool intended to organize and curate links to information about data sets relevant to our 3D understanding of the Milky Way. For any given longitude range, we provide the means to determine the available surveys, their overlapping footprint, and the type of data each provides. Information about each dataset, including how to access the data, their hallmark publications, and their principal investigators, is available at the Milky Way 3D Dataverse. All the data can be loaded, "linked", and explored using the new 3D visualization software package Glue, available for download at glueviz.org!



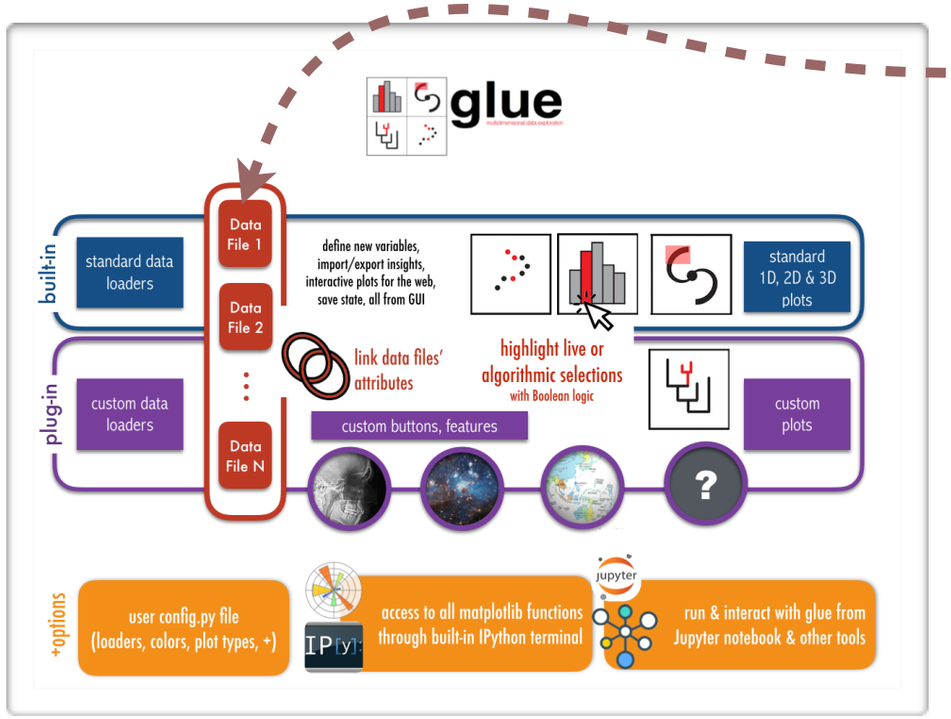
This is an interactive coverage tool of surveys of the galactic plane created with the AAS WorldWide Telescope API, and WorldWide Telescope. To view the coverage of a region, click on the check box or slider in the first column of the table below. You can drag your mouse to pan or zoom in the window. Information on where to find the data in the regions may be found by clicking on the region name or the ★. Table columns may be sorted by clicking on the table header.

Zoom Out | Zoom In | Reset/Center

			Extended Observations		Catalogs and Pointed Surveys	
View Region	Link to Survey	Wavelength	Continuum (2D)	Spectral Line (3D)	Source-Based Lists	Spectral Line
☑	THOR	21 cm, 300 mm, 174-186 mm		★		
☑	BESSEL	1-3 cm			★	
☑	HOPS*	1 cm		★		
☑	CORNISH*	60 mm	★		★	
☑	HOPS	12 mm		★		★
☑	GRS	3 mm		★		
☑	MALT90	3 mm				★
☑	THRUMMS	3 mm		★		
☑	Dame CO	2.6 mm		★		
☑	BGPS	1 mm	★		★	★
☑	CHIMPS	1 mm		★		
☑	COHRS	1 mm		★		
☑	ATLASGAL	870 μm	★		★	
☑	JCMT*	850 μm	★		★	
☑	HIGAL*	70-500 μm	★			
☑	MIPSGAL	24, 70 μm	★			
☑	WISE	3.4, 4.6, 12, 22.0 μm	★			
☑	GLIMPSE	3.6, 4.5, 5.8, 8.0 μm	★			
☑	UKIDSS-GPS*	1.3, 1.6, 2.2 μm	★			
☑	GPIPS	1.6 μm			★	

☑ Toggle All

*Future data to be released



The Dataverse Project

About | Guides | Support | Sign Up | Log In

The HI, OH, Recombination Line Survey of the Milky Way (THOR) Survey Dataverse (Harvard-Smithsonian Center for Astrophysics)

Harvard Dataverse > MilkyWay3D Dataverse > **The HI, OH, Recombination Line Survey of the Milky Way (THOR) Survey Dataverse**

The THOR survey is a galactic plane survey covering 15° < l < 67° and b < |1°|. It consists of HI 21 cm observations, 20 recombination lines and 1-2 GHz continuum data at a 20" spatial resolution. One of its main science goals is understanding the dynamics of the interstellar medium and how clouds transition from the atomic to the molecular phase. The link to the homepage of the THOR survey can be found here.

Search this dataverse... | Find | Advanced Search

- Dataverses (0)
- **Datasets (1)**
- Files (0)

Publication Date
2016 (1)

Subject
Astronomy and Astrophysics (1)

Author Name
Simon Bihr (1)

1 to 1 of 1 Result | Sort

THOR Extended Observations: Spectral Line (3D)
Sep 5, 2016
Simon Bihr, 2016, "THOR Extended Observations: Spectral Line (3D)", doi:10.7910/DVN/1ZJGAO, Harvard Dataverse, V1
This dataset consists of extended spectral line observations taken by the Very Large Array (VLA), over l = 15° – 67°, |b| ≤ 1°. There are a variety of data products available. The first includes HI 21 cm observations with a bandwidth of 2 MHz and a channel width of 1.953 kHz. Thi...

Developed at the Institute for Quantitative Social Science | Dataverse Project on | Code available at

Copyright © 2016, The President & Fellows of Harvard College | Privacy Policy

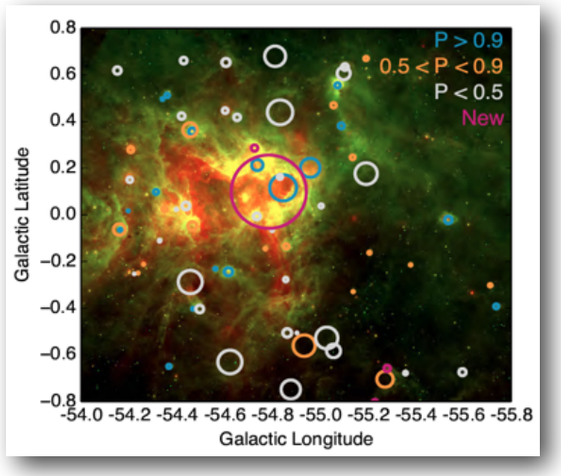
Powered by The Dataverse Project v. 4.5 build 41-dc58ae1

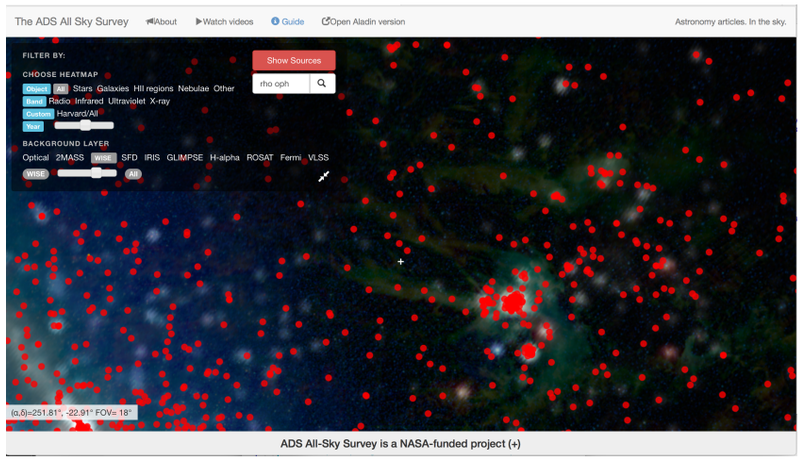
(remote access to) Big (and Wide) Data

Bridges to Web (e.g. Jupyter Labs, bokeh)

(more) publishing options

More?





ADS All-Sky Survey is a NASA-funded project (+)

10 ViZ

AUGUST 22-25, 2017
NEW YORK, NY

jupytercon.com
#JupyterCon

Tools
RStudio
H2O
learn
jupyter
plotly
knitr

Publishing & Deployment

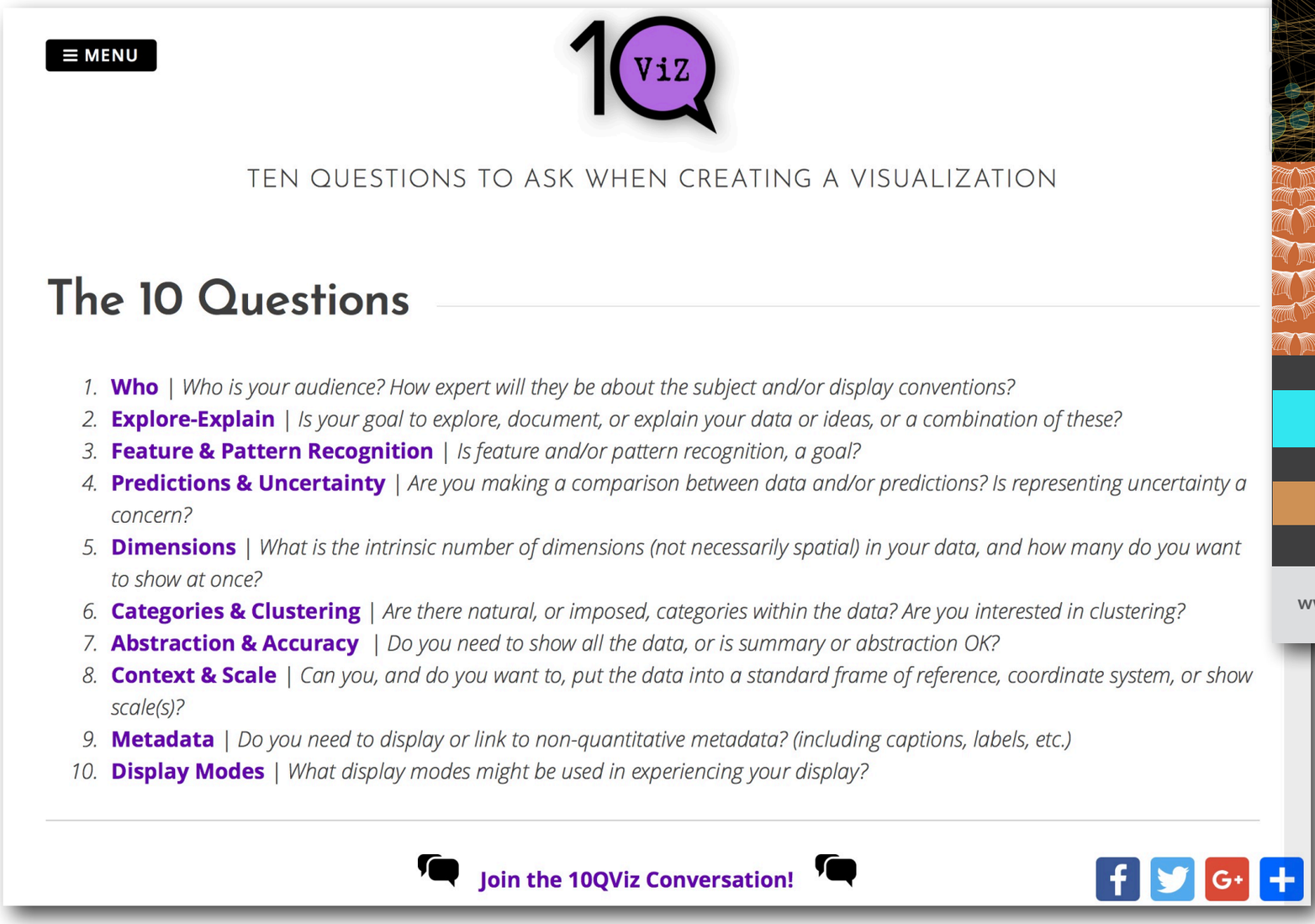
into the Business

More Time for Data Science

©2017 Domino Data Lab

DOMINO

To continue the conversation…



10qviz.org *with Arzu Çöltekin (beta 2017, release 2018)*

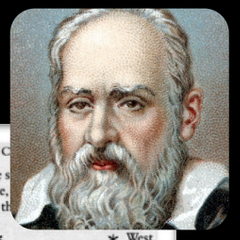
Creativity & Collaboration at NAS March 2018

with Ben Shneiderman, Maneesh Agrawala, Roger Malina, Youngmoo Kim & Donna Cox

(more) publishing options

4 Centuries from Galileo to Galileo

1610



SIDEREUS NUNC

On the third, at the seventh hour, the s
sequence. The eastern one was 1 minute,
the closest western one 2 minutes; and th

East * ○ * * West

10 minutes removed from this one. They were absolutely on the same straight line and of equal magnitude.

On the fourth, at the second hour, there were four stars around Jupiter, two to the east and two to the west, and arranged precisely

East * * ○ * * West

on a straight line, as in the adjoining figure. The easternmost was distant 3 minutes from the next one, while this one was 40 seconds from Jupiter; Jupiter was 4 minutes from the nearest western one, and this one 6 minutes from the westernmost one. Their magnitudes were nearly equal; the one closest to Jupiter appeared a little smaller than the rest. But at the seventh hour the eastern st
30 seconds apart. Jupiter was 2 minutes from the

East ** ○ * *

one, while he was 4 minutes from the next western
one was 3 minutes from the westernmost one. They
and extended on the same straight line along the ecl

On the fifth, the sky was cloudy.

On the sixth, only two stars appeared flanking Ju

East * ○ *

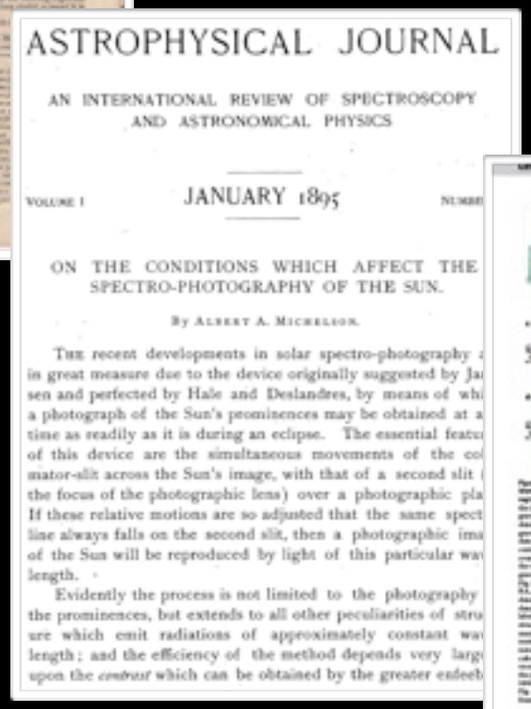
in the adjoining figure. The eastern one was 2 mi
western one 3 minutes from Jupiter. They were on th
line with Jupiter and equal in magnitude.

On the seventh, two stars stood near Jupiter, bo
arranged in this manner.

1665

PHILOSOPHICAL
TRANSACTIONS

1895



ASTROPHYSICAL JOURNAL

AN INTERNATIONAL REVIEW OF SPECTROSCOPY AND ASTRONOMICAL PHYSICS

VOLUME I JANUARY 1895 NUMBE

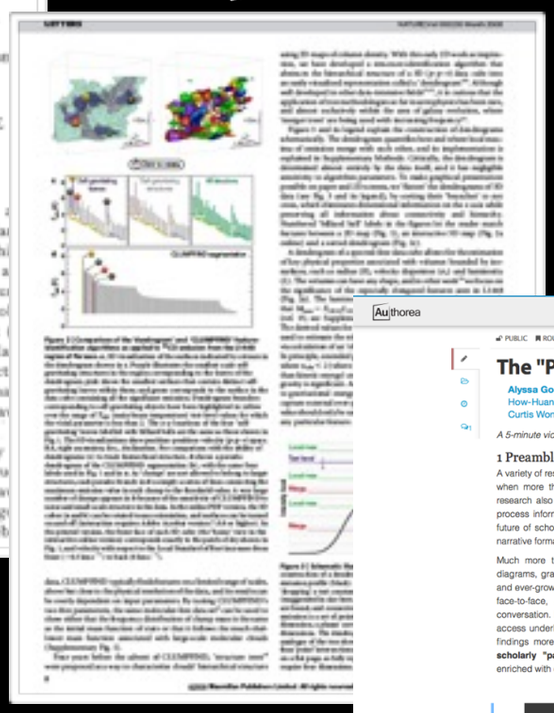
ON THE CONDITIONS WHICH AFFECT THE SPECTRO-PHOTOGRAPHY OF THE SUN.

By Albert A. Michelson.

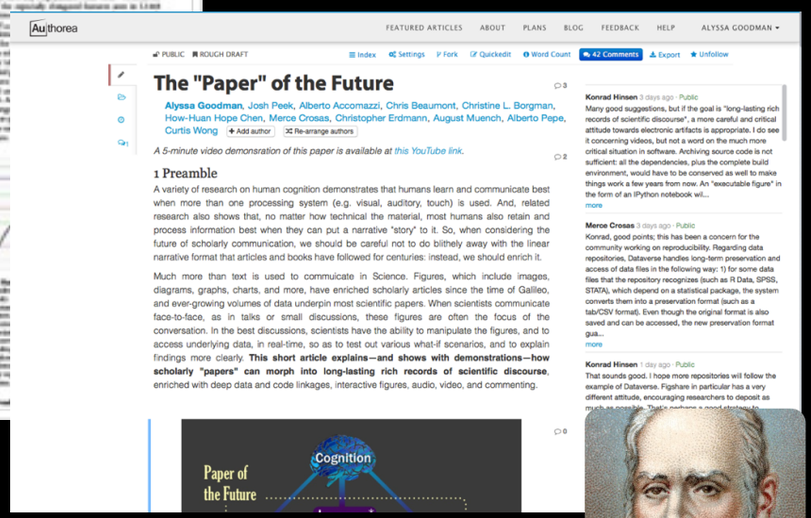
The recent developments in solar spectro-photography
in great measure due to the device originally suggested by Ja
sen and perfected by Hale and Deslandres, by means of wh
a photograph of the Sun's prominences may be obtained at a
time as readily as it is during an eclipse. The essential featu
of this device are the simultaneous movements of the co
mator-slit across the Sun's image, with that of a second slit
the focus of the photographic lens) over a photographic pla
If these relative motions are so adjusted that the same spect
line always falls on the second slit, then a photographic ima
of the Sun will be reproduced by light of this particular wa
length.

Evidently the process is not limited to the photography
the prominences, but extends to all other peculiarities of stru
ure which emit radiations of approximately constant wa
length; and the efficiency of the method depends very larg
upon the *contrast* which can be obtained by the greater endeeb

2009



2015



Authorea

FEATURED ARTICLES ABOUT PLANS BLOG FEEDBACK HELP ALYSSA GOODMAN

PUBLIC ROUGH DRAFT

Index Settings Fork Quickedit Word Count 42 Comments Export Unfollow

The "Paper" of the Future

Alyssa Goodman, Josh Peek, Alberto Accomazzi, Chris Beaumont, Christine L. Borgman, How-Huan Hope Chen, Merce Crosas, Christopher Erdmann, August Muench, Alberto Pepe, Curtis Wong + Add author Re-arrange authors

A 5-minute video demonsration of this paper is available at this YouTube link.

1 Preamble

A variety of research on human cognition demonstrates that humans learn and communicate best when more than one processing system (e.g. visual, auditory, touch) is used. And, related research also shows that, no matter how technical the material, most humans also retain and process information best when they can put a narrative "story" to it. So, when considering the future of scholarly communication, we should be careful not to do blithely away with the linear narrative format that articles and books have followed for centuries: instead, we should enrich it.

Much more than text is used to commuicate in Science. Figures, which include images, diagrams, graphs, charts, and more, have enriched scholarly articles since the time of Galileo, and ever-growing volumes of data underpin most scientific papers. When scientists communicate face-to-face, as in talks or small discussions, these figures are often the focus of the conversation. In the best discussions, scientists have the ability to manipulate the figures, and to access underlying data, in real-time, so as to test out various what-if scenarios, and to explain findings more clearly. **This short article explains—and shows with demonstrations—how scholarly "papers" can morph into long-lasting rich records of scientific discourse**, enriched with deep data and code linkages, interactive figures, audio, video, and commenting.

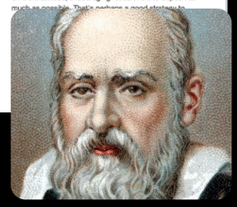
Paper of the Future

Cognition

Konrad Hinsen 3 days ago · Public
Many good suggestions, but if the goal is "long-lasting rich records of scientific discourse", a more careful and critical attitude towards electronic artifacts is appropriate. I do see it concerning videos, but not a word on the much more critical situation in software. Archiving source code is not sufficient: all the dependencies, plus the complete build environment, would have to be conserved as well to make things work a few years from now. An "executable figure" in the form of an IPython notebook wi...
more

Merce Crosas 3 days ago · Public
Konrad, good points; this has been a concern for the community working on reproducibility. Regarding data repositories, Dataverse handles long-term preservation and access of data files in the following way: 1) for some data files that the repository recognizes (such as R Data, SPSS, STATA), which depend on a statistical package, the system converts them into a preservation format (such as a tab/CSV format). Even though the original format is also saved and can be accessed, the new preservation format gua...
more

Konrad Hinsen 1 day ago · Public
That sounds good. I hope more repositories will follow the example of Dataverse. Figshare in particular has a very different attitude, encouraging researchers to deposit as



glue
multidimensional data exploration

enabled by d3.js (javascript) outputs

plotly

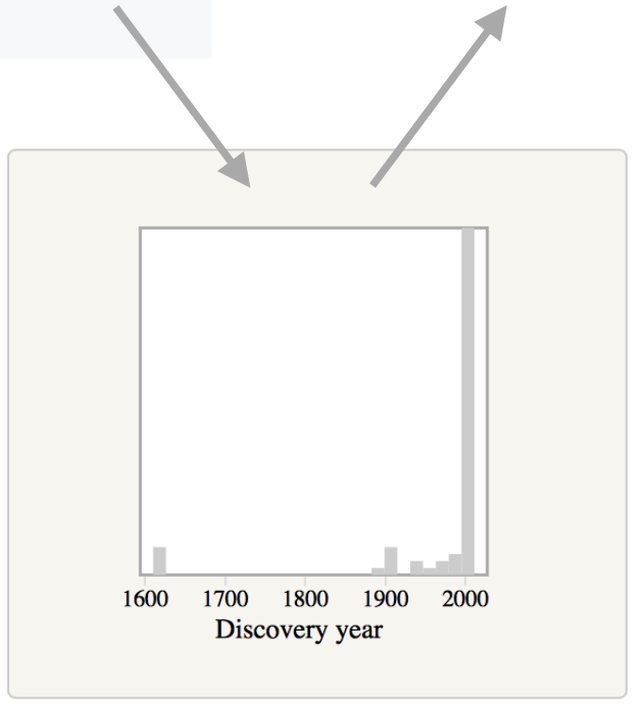
d3po

d3po is a project designed to allow an astronomer (or anyone), with no special data visualization skills, to make an interactive, publication-quality figure that has staged builds and linked brushing through scatter plots. Our current version can be previewed at d3po.org, and represents a figure from upcoming work by graduate student Elisabeth Newton. The figure describes how metalicity affects color in cool stars, and represents a nice use case for d3po. Try clicking and dragging in the scatter plots to understand the power of linked brushing in published figures.

Right now we are in search of alpha testers, who have figures that could be made interactive and who are willing to get their hands a little dirty (No javascript skills needed). In future versions, we plan to link to glue to allow the creation of d3po figures interactively. We are also exploring implementation of d3po within presentations and within authorea. Full 1.0 version expected in January 2014.

Installing your own d3po server

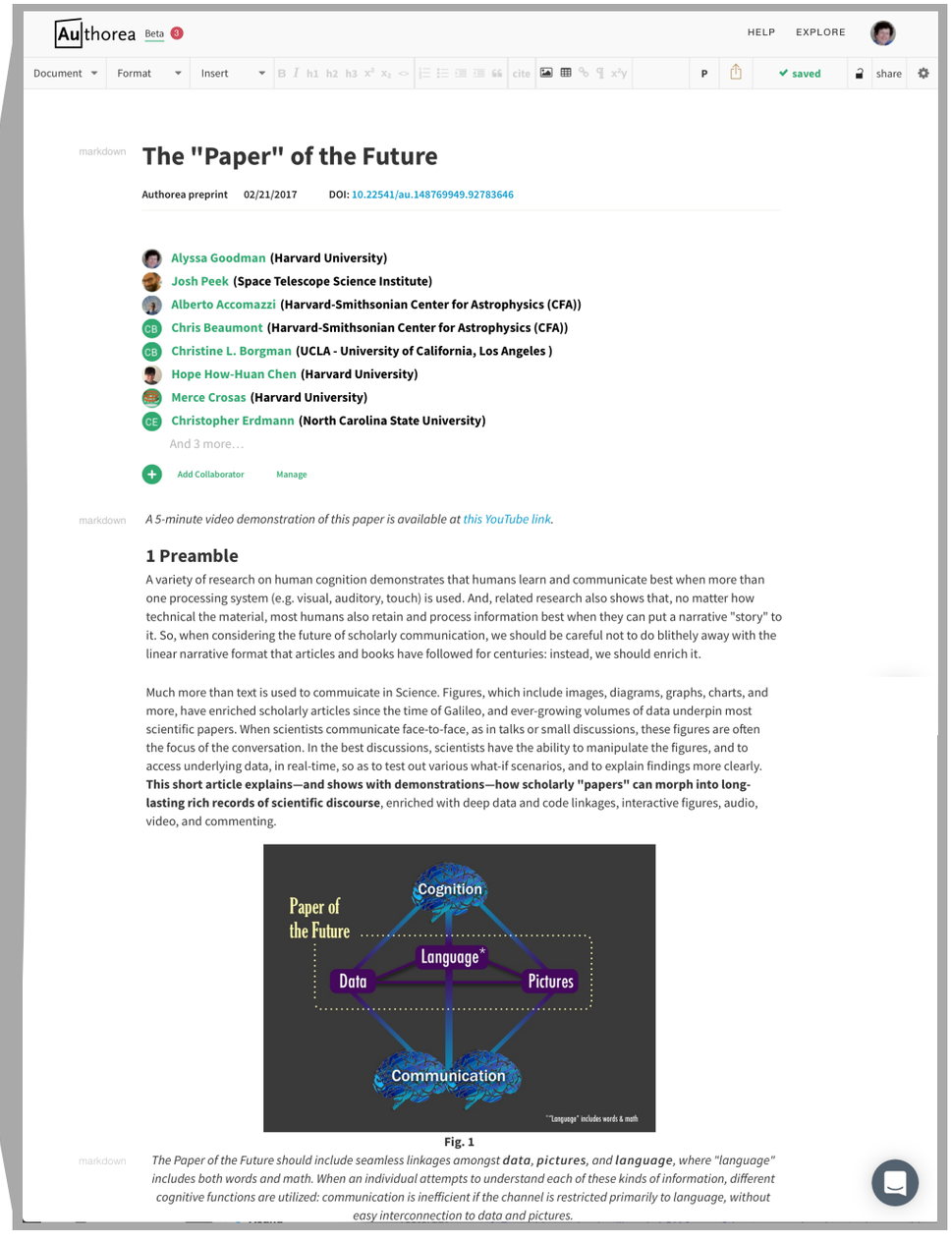
```
git clone git@github.com:adrn/d3po.git
cd d3po
virtualenv --no-site-packages venv
source venv/bin/activate
pip install -r pip-requirements.txt
python run.py
```



Four Centuries of Discovery | A Chasm in Mass | Little Siblings | Close Cousins | The Strangers

After Galileo discovered the first four moons of Jupiter, it took nearly three hundred years to discover the next one.

Authorea



[demo]

Many thanks to Alberto Pepe, Josh Peek, Chris Beaumont, Tom Robitaille, Adrian Price-Whelan, Elizabeth Newton, Michelle Borkin & Matteo Cantiello for making this posible.

2009

3D PDF

High-Dimensional data in a "Paper" on its way to the Future

[demo/video]

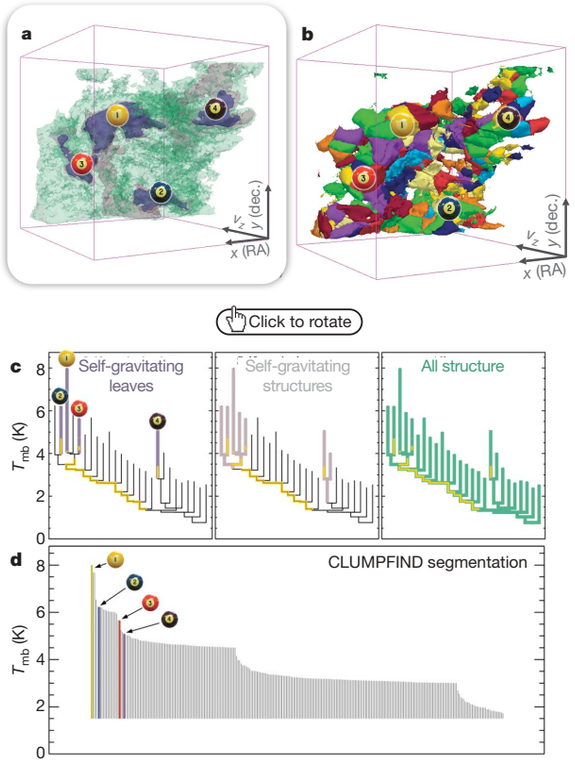


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating structures in the region corresponding to the leaves of the dendrogram; pink shows the smallest surfaces that contain distinct self-gravitating leaves within them; and green corresponds to the surface in the data cube containing all the significant emission. Dendrogram branches corresponding to self-gravitating objects have been highlighted in yellow over the range of T_{mb} (main-beam temperature) test-level values for which the virial parameter is less than 2. The x–y locations of the four 'self-gravitating' leaves labelled with billiard balls are the same as those shown in Fig. 1. The 3D visualizations show position–position–velocity (p–p–v) space. RA, right ascension; dec., declination. For comparison with the ability of dendrograms (**c**) to track hierarchical structure, **d** shows a pseudo-dendrogram of the CLUMPFIND segmentation (**b**), with the same four labels used in Fig. 1 and in **a**. As 'clumps' are not allowed to belong to larger structures, each pseudo-branch in **d** is simply a series of lines connecting the maximum emission value in each clump to the threshold value. A very large number of clumps appears in **b** because of the sensitivity of CLUMPFIND to noise and small-scale structure in the data. In the online PDF version, the 3D cubes (**a** and **b**) can be rotated to any orientation, and surfaces can be turned on and off (interaction requires Adobe Acrobat version 7.0.8 or higher). In the printed version, the front face of each 3D cube (the 'home' view in the interactive online version) corresponds exactly to the patch of sky shown in Fig. 1, and velocity with respect to the Local Standard of Rest increases from front (-0.5 km s^{-1}) to back (8 km s^{-1}).

data, CLUMPFIND typically finds features on a limited range of scales, above but close to the physical resolution of the data, and its results can be overly dependent on input parameters. By tuning CLUMPFIND's two free parameters, the same molecular-line data set[8] can be used to show either that the frequency distribution of clump mass is the same as the initial mass function of stars or that it follows the much shallower mass function associated with large-scale molecular clouds (Supplementary Fig. 1).

Four years before the advent of CLUMPFIND, 'structure trees'[9] were proposed as a way to characterize clouds' hierarchical structure using 2D maps of column density. With th… …ly 2D work as inspiration, we have developed a structure-id… …ithm that abstracts the hierarchical structure of a… an easily visualized representation calle… well developed in other data-intensive … application of tree methodologies so fa… and almost exclusively within the a… 'merger trees' are being used with in…

Figure 3 and its legend explain th… schematically. The dendrogram qua… ima of emission merge with each… explained in Supplementary Meth… determined almost entirely by th… sensitivity to algorithm paramet… possible on paper and 2D screen… data (see Fig. 3 and its legend… cross, which eliminates dimens… preserving all information… Numbered 'billiard ball' lab… features between a 2D map… online) and a sorted dendro…

A dendrogram of a spectr… of key physical properties… surfaces, such as radius (R), ve… (L). The volumes can have any shape, and n… the significance of the especially elongated features … (Fig. 2a). The luminosity is an approximate proxy for mass, such that $M_{lum} = X_{13CO} L_{13CO}$, where $X_{13CO} = 8.0 \times 10^{20}$ cm^2 K^{-1} km^{-1} s (ref. 15; see Supplementary Methods and Supplementary Fig. 2). The derived values for size, mass and velocity dispersion can then be used to estimate the role of self-gravity at each point in the hierarchy, via calculation of an 'observed' virial parameter, $\alpha_{obs} = 5\sigma_v^2 R/GM_{lum}$. In principle, extended portions of the tree (Fig. 2, yellow highlighting) where $\alpha_{obs} < 2$ (where gravitational energy is comparable to or larger than kinetic energy) correspond to regions of p–p–v space where self-gravity is significant. As α_{obs} only represents the ratio of kinetic energy to gravitational energy at one point in time, and does not explicitly capture external over-pressure and/or magnetic fields[16], its measured value should only be used as a guide to the longevity (boundedness) of any particular feature.

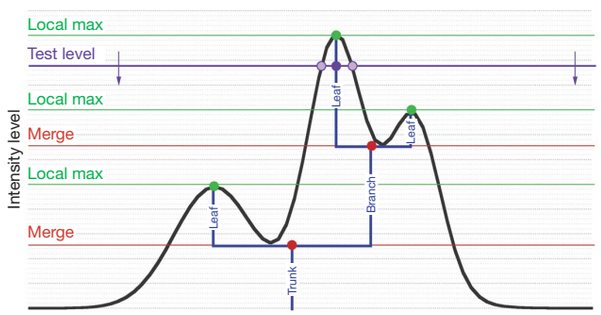
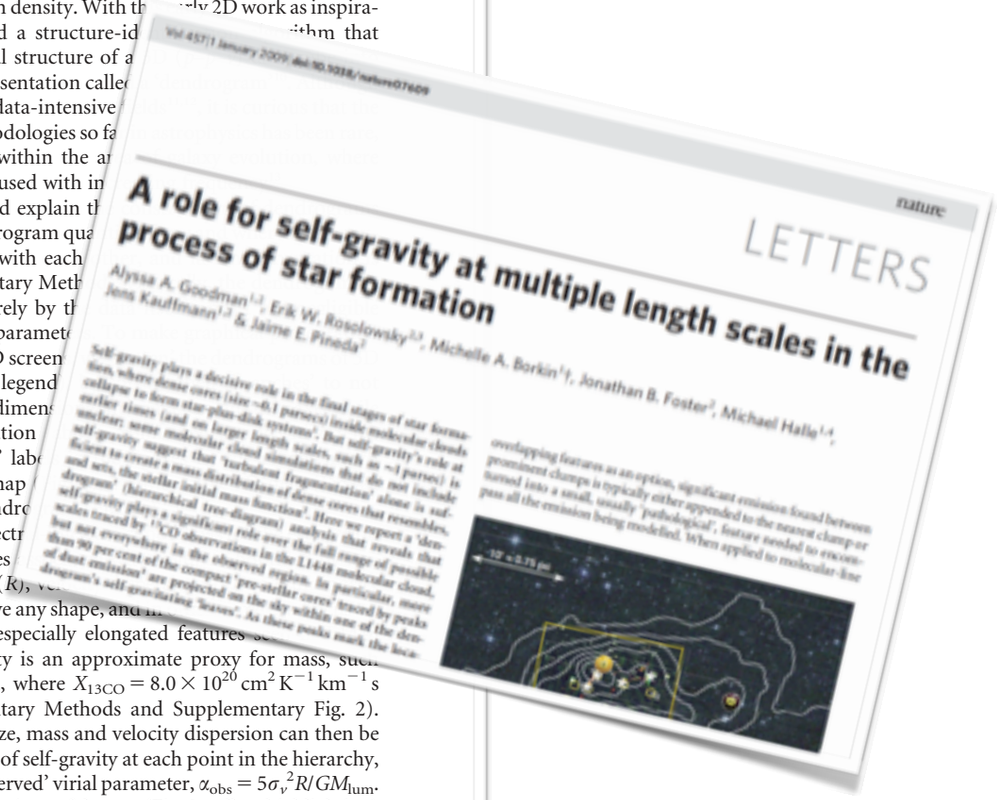
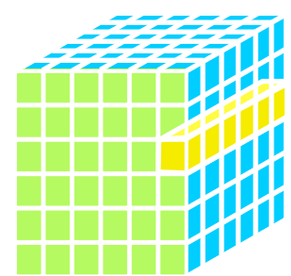


Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in size here, light lines) until all the local maxima and mergers are found, and connected as shown. The intersection of a test level with the emission is a set of points (for example the light purple dots) in one dimension, a planar curve in two dimensions, and an isosurface in three dimensions. The dendrogram of 3D data shown in Fig. 2c is the direct analogue of the tree shown here, only constructed from 'isosurface' rather than 'point' intersections. It has been sorted and flattened for representation on a flat page, as fully representing dendrograms for 3D data cubes would require four dimensions.

©2009 Macmillan Publishers Limited. All rights reserved





Goodman et al. 2009, Nature, cf: Fluke et al. 2009

Vol 457|1 January 2009|doi:10.1038/nature07609 nature

LETTERS

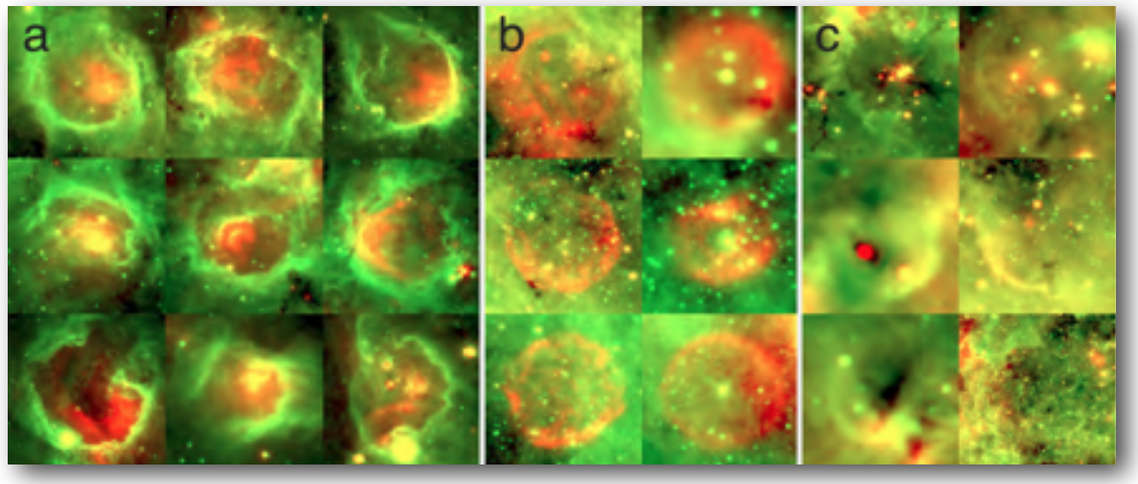
A role for self-gravity at multiple length scales in the process of star formation

Alyssa A. Goodman[1,2], Erik W. Rosolowsky[2,3], Michelle A. Borkin[1]†, Jonathan B. Foster[2], Michael Halle[1,4], Jens Kauffmann[1,2] & Jaime E. Pineda[2]

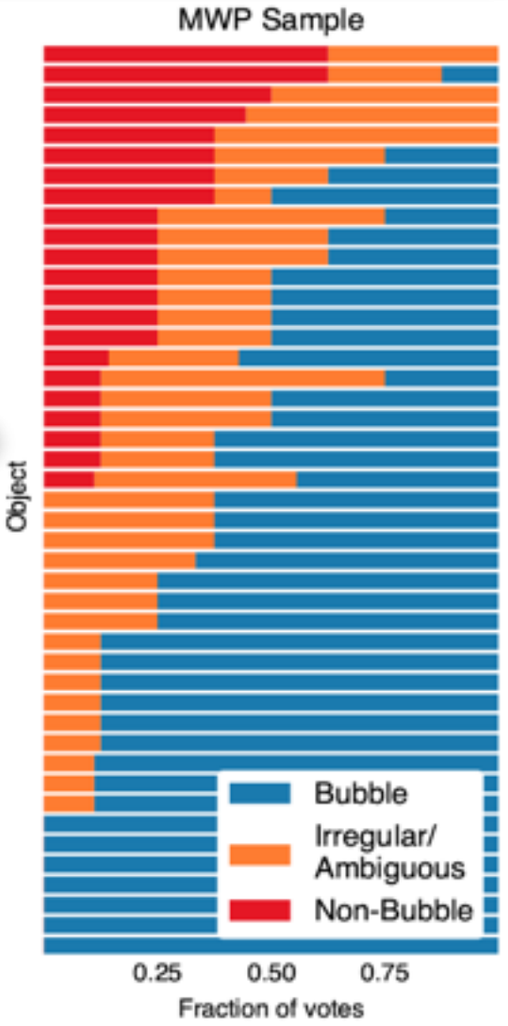
Self-gravity plays a decisive role in the final stages of star formation, where dense cores (size ~0.1 parsecs) inside molecular clouds collapse to form star-plus-disk systems[1]. But self-gravity's role at earlier times (and on larger length scales, such as ~1 parsec) is unclear; some molecular cloud simulations that do not include self-gravity suggest that 'turbulent fragmentation' alone is sufficient to create a mass distribution of dense cores that resembles, and sets, the stellar initial mass function[2]. Here we report a 'dendrogram' (hierarchical tree-diagram) analysis that reveals that self-gravity plays a significant role over the full range of possible scales traced by ^{13}CO observations in the L1448 molecular cloud, but not everywhere in the observed region. In particular, more than 90 per cent of the compact 'pre-stellar cores' traced by peaks of dust emission[3] are projected on the sky within one of the dendrogram's self-gravitating 'leaves'. As these peaks mark the locations of already-forming stars, or of those probably about to form, a self-gravitating cocoon seems a critical condition for their exist-

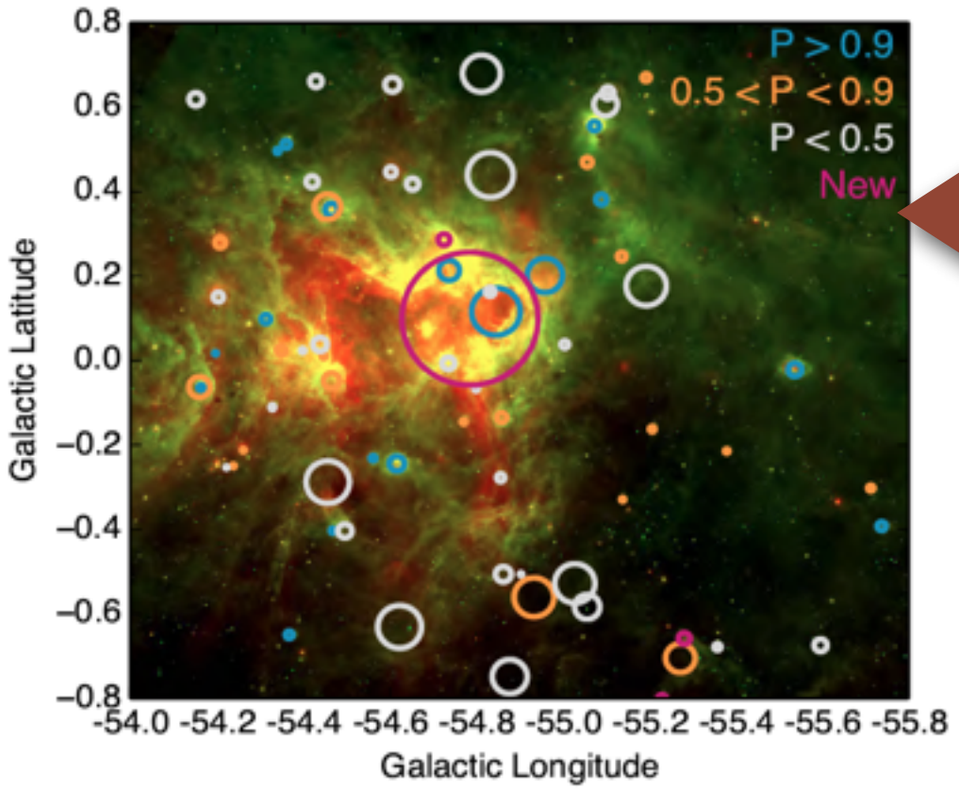
overlapping features as an option, significant emission found between prominent clumps is typically either appended to the nearest clump or turned into a small, usually 'pathological', feature needed to encompass all the emission being modelled. When applied to molecular-line

BIG DATA AND "HUMAN-AIDED COMPUTING"



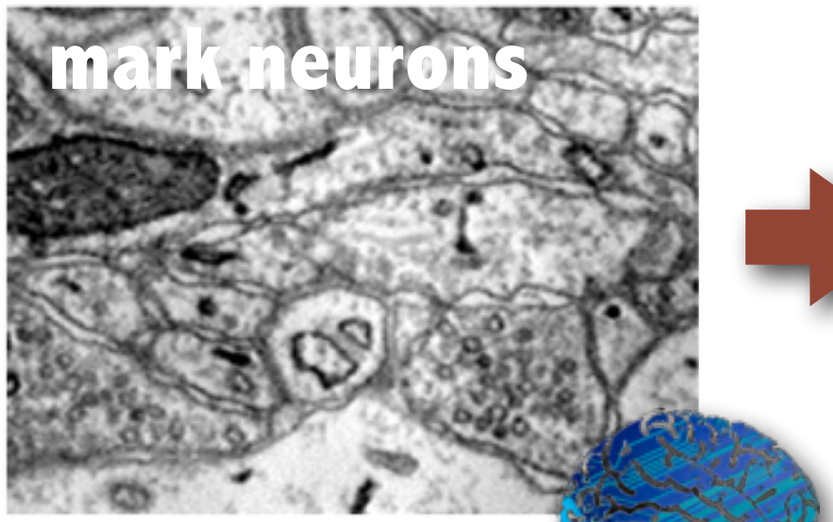
machine-learning algorithm (Brut)

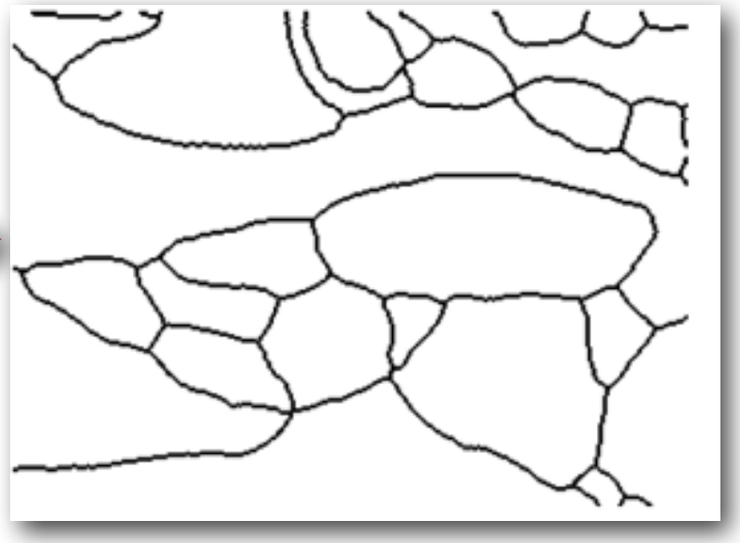




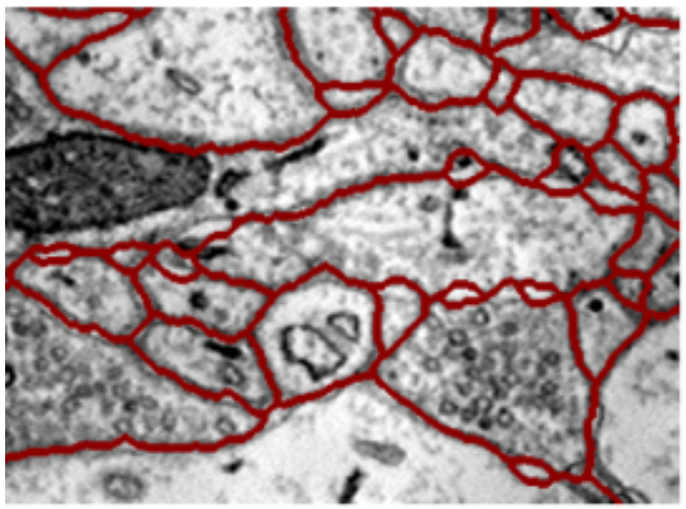
example here from: ***Beaumont****, Goodman, Kendrew, Williams & Simpson 2014; based on* ***Milky Way Project*** *catalog (Simpson et al. 2013), which came from* ***Spitzer/GLIMPSE*** *(Churchwell et al. 2009, Benjamin et al. 2003), cf. Shenoy & Tan 2008 for discussion of HAC;* ***astroml.org*** *for machine learning advice/tools*

BIG DATA AND "HUMAN-AIDED COMPUTING"



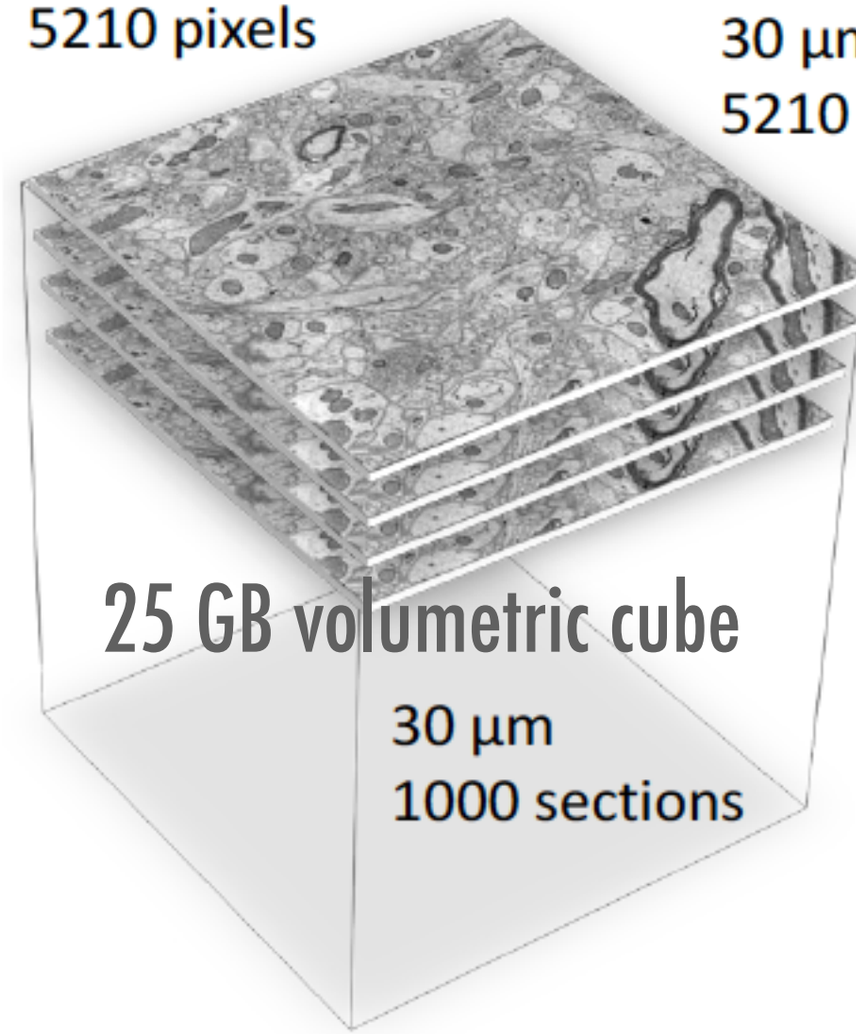


machine-learning algorithm (RF+CRF)

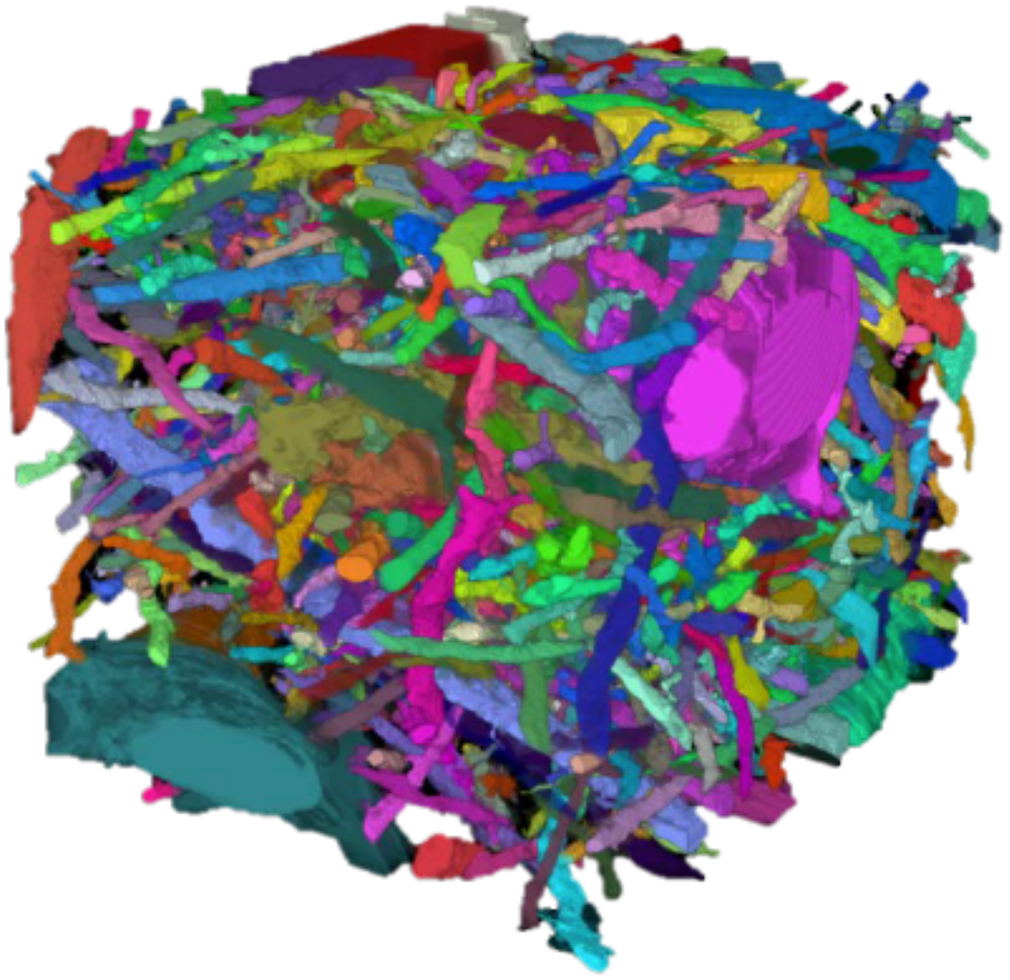


example here from: Kaynig...Lichtman...Pfister et al. 2013, "Large-Scale Automatic Reconstruction of Neuronal Processes from Electron Microscopy Images"; cf. Shenoy & Tan 2008 for discussion of HAC; ***astroml.org*** *for machine learning advice/tools (Note: RF=Random Forest; CRF=Conditional Random Fields.)*

BIG DATA AND "HUMAN-AIDED COMPUTING"



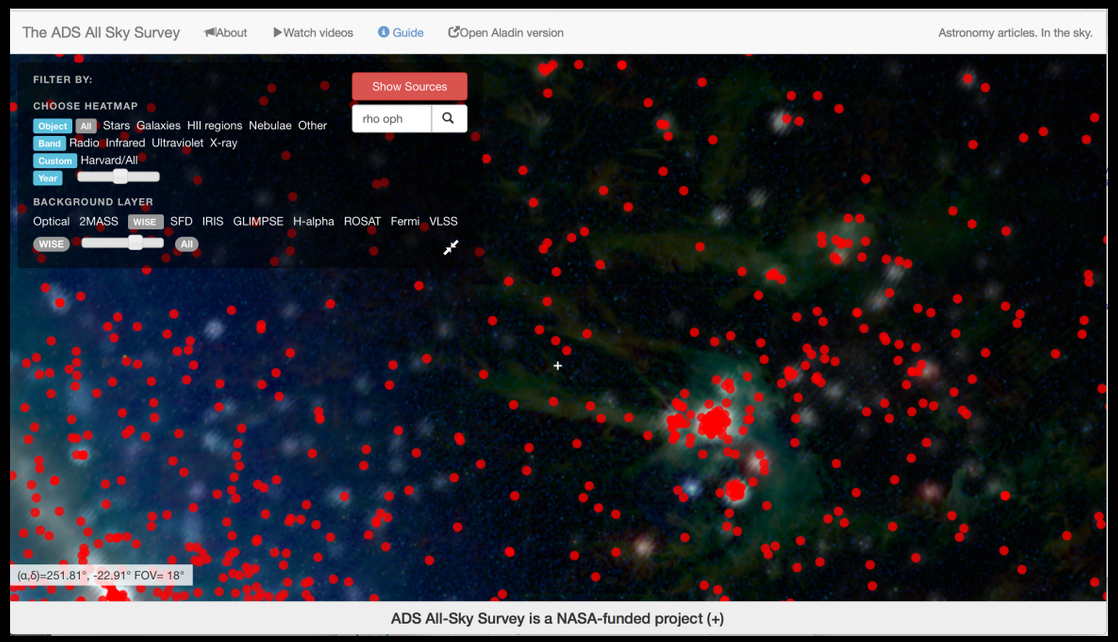
EM images

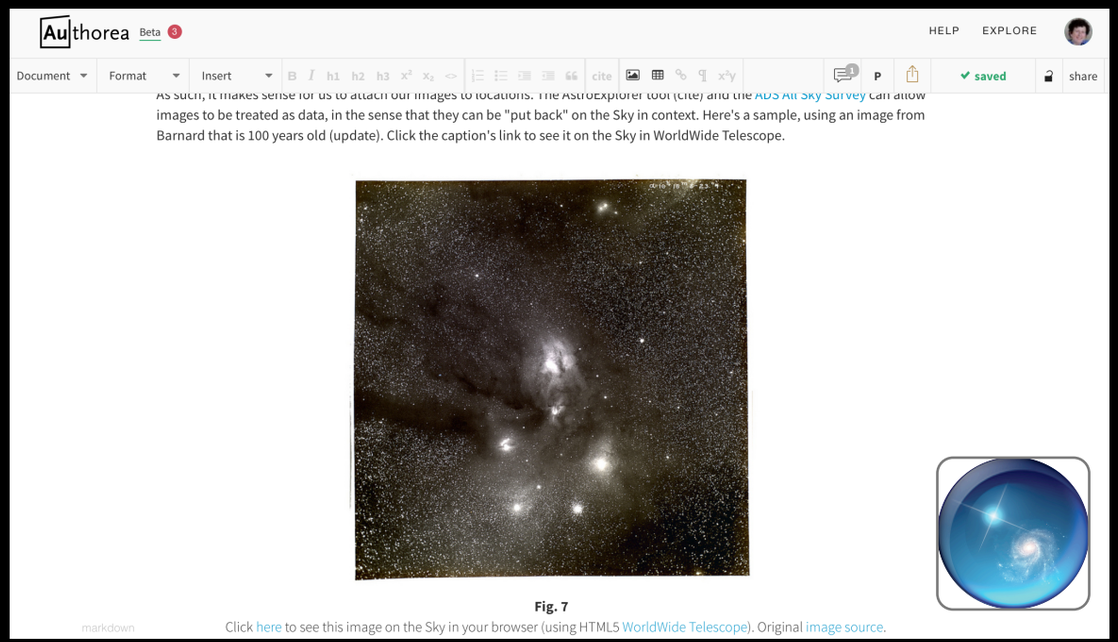


Reconstruction

example here from: Kaynig...Lichtman...Pfister et al. 2013, "Large-Scale Automatic Reconstruction of Neuronal Processes from Electron Microscopy Images"; cf. Shenoy & Tan 2008 for discussion of HAC; ***astroml.org*** *for machine learning advice/tools*

Literature as (a filter for) Data





Many thanks to Alberto Pepe, August Muench, Thomas Boch, Jonathan Fay, Michael Kurtz, Alberto Accomazzi, Julie Steffen, Laura Trouille, David Hogg, Dustin Lang, Christopher Stumm, Chris Beaumont & Phil Rosenfield for making this all work!

ADS All-Sky Survey & Astronomy Rewind

"putting articles and images (back) on the Sky"

1. Images Extracted from Journal Articles

ON A GREAT NEBULOUS REGION AND ON THE QUESTION OF ABSORBING MATTER IN SPACE AND THE TRANSPARENCY OF THE NEBULAE

By E. E. BARNARD

While photographing the region of the great nebula of *ρ Ophiuchi* (which I had found with the Willard lens) at the Lick Observatory in 1893, the plates with the small lantern lens (1½ inches diameter, also attached to the Willard mounting) showed a remarkable nebula involving the 4.5 magnitude star *ν Scorpii* (Plate I). It had not been noticed on the Willard lens photograph, where it was very faint and near the edge of the plate. The discovery of this object therefore is due to the small lantern lens.

astrophysics data system

"historical" images

THE COMPLETE SURVEY OF STAR-FORMING REGIONS: PHASE I DATA

Naomi A. Ridge,[1] James Di Francisco,[2] Helen Kirk,[2,3] Di Li,[1,4] Alyssa A. Goodman,[1] João F. Alves,[5] Héctor G. Arce,[6] Michelle A. Borkin,[7] Paola Caselli,[8] Jonathan B. Foster,[1] Mark H. Heyer,[9] Doug Johnstone,[2,3] David A. Kosslyn,[1] Marco Lombardi,[5] Jaime E. Pineda,[1] Scott L. Scibelli,[1] and Mario Tafalla[10]

Received 2005 November 8; accepted 2006 February 22

ABSTRACT

We present an overview of data available for the Ophiuchus and Perseus molecular clouds from Phase I of the COMPLETE Survey of Star-Forming Regions. This survey provides a range of data complementary to the *Spitzer*

astrophysics data system

astronomy image explorer

"recent" images

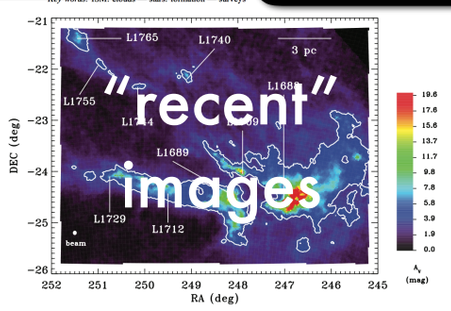


Fig. 3.—Map of extinction in Ophiuchus derived using 2MASS NICER. The contour indicates an A_V of 3 mag and is repeated in subsequent figures for orientation. Note that the small "hole" at the center of the L1688 cluster is an artifact due to the high extinction at that position.

2. Missing coordinate metadata added back to images, either...

...automatically, applying astronometry.net to wide-field optical images, or

Astrometry.net

via "Astronomy Rewind" Zooniverse Citizen Science Project

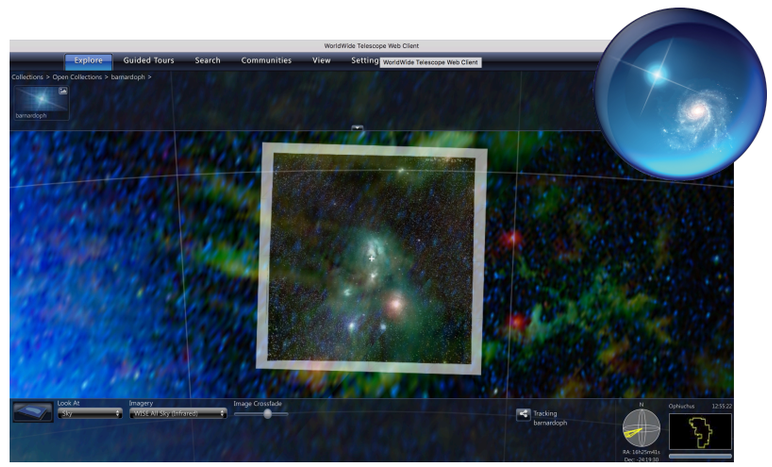
ZOONIVERSE

3. "Solved" images returned to ADS & Astronomy Image Explorer

astrophysics data system

astronomy image explorer

4. New button in Astronomy Image Explorer offers image-in-context, using AAS' WorldWide Telescope in the browser



your handout

0. ADS All-Sky Survey ticle density layer on the Sky

ADS ALL SKY SURVEY

click entries on the timeline to try out services

- **ADS** 1992
- **WorldWide Telescope** 2008
- **Zooniverse** 2009
- **Astrometry.net** 2011
- **ADS All Sky Survey** 2014
- **Astronomy Image Explorer** 2014
- **Astronomy Rewind** 2017

The Astronomical Journal, 131:2921–2933, 2006 June
© 2006. The American Astronomical Society. All rights reserved. Printed in U.S.A.

THE COMPLETE SURVEY OF STAR-FORMING REGIONS: PHASE I DATA

Naomi A. Ridge,[1] James Di Francesco,[2] Helen Kirk,[2,3] Di Li,[1,4] Alyssa A. Goodman,[1] João F. Alves,[5] Héctor G. Arce,[6] Michelle A. Borkin,[7] Paola Caselli,[8] Jonathan B. Foster,[1] Mark H. Heyer,[9] Doug Johnstone,[2,3] David A. Kosslyn,[1] Marco Lombardi,[4] Jaime E. Pineda,[1] Scott L. Schnee,[1] and Mario Tafalla[10]

Received 2005 November 8; accepted 2006 February 22

ABSTRACT

We present an overview of data available for the Ophiuchus and Perseus molecular clouds from Phase I of the COMPLETE Survey of Star-Forming Regions. This survey provides a range of data complementary to the *Spitzer* Legacy Program "From Molecular Cores to Planet Forming Disks." Phase I includes the following: extinction maps derived from the Two Micron All Sky Survey (2MASS) near-infrared data using the NICER algorithm; extinction and temperature maps derived from *IRAS* 60 and 100 μm emission; H I maps of atomic gas; ^{12}CO and ^{13}CO maps of molecular gas; and submillimeter continuum images of emission from dust in dense cores. Not unexpectedly, the morphology of the regions appears quite different depending on the column density tracer that is used, with *IRAS* tracing mainly warmer dust and CO being biased by chemical, excitation, and optical depth effects. Histograms of column density distribution are presented, showing that extinction as derived from 2MASS NICER gives the closest match to a lognormal distribution, as is predicted by numerical simulations. All the data presented in this paper, and links to more detailed publications on their implications, are publicly available at the COMPLETE Web site.

Key words: ISM: clouds — stars: formation — surveys

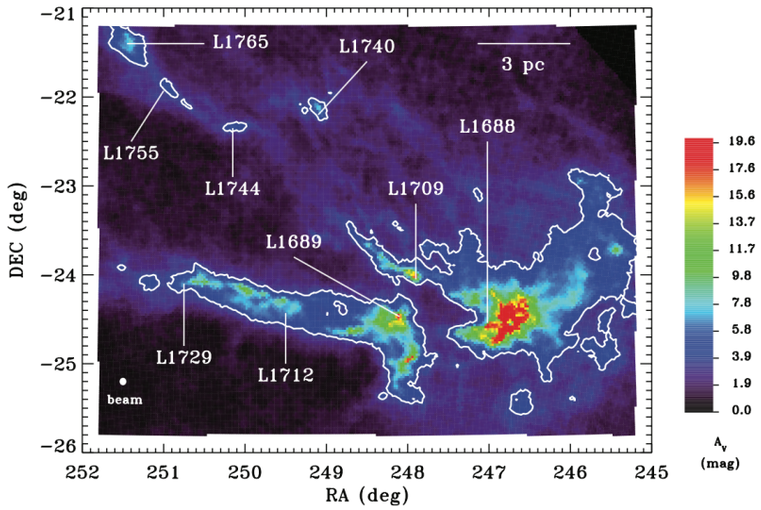
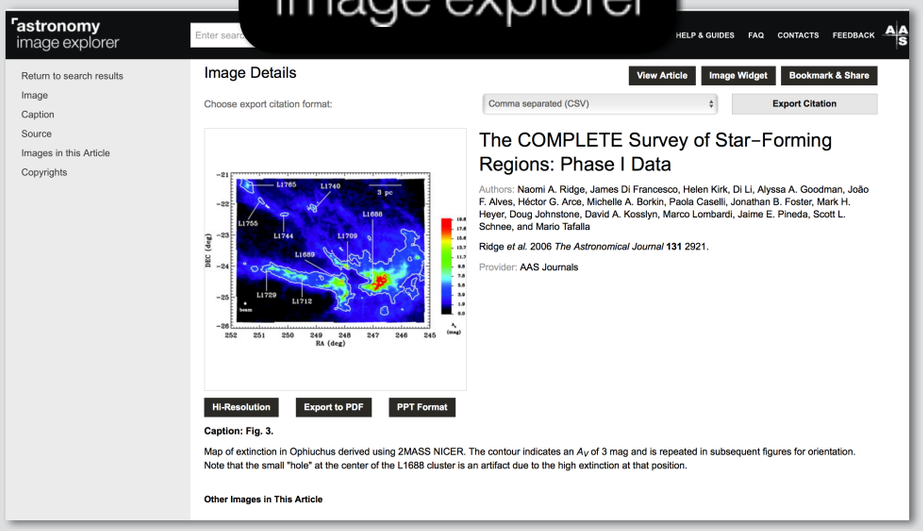


Fig. 3.—Map of extinction in Ophiuchus derived using 2MASS NICER. The contour indicates an A_V of 3 mag and is repeated in subsequent figures for orientation. Note that the small "hole" at the center of the L1688 cluster is an artifact due to the high extinction at that position.

astronomy image explorer



Who, How, and Who's Paying?

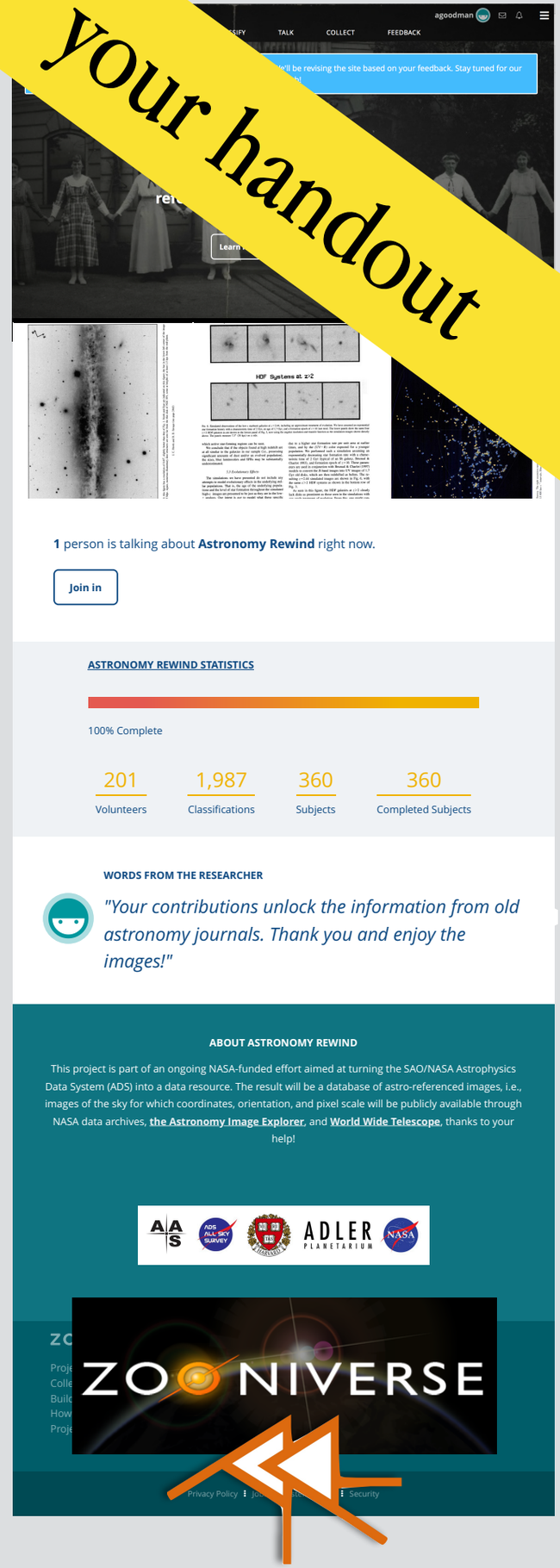
The **ADS All Sky Survey** was first funded via a 2012 grant from the **NASA ADAP** program to Seamless Astronomy, in collaboration with CDS, Astrometry.net and Microsoft Research.

Articles-on-the-Sky was first deployed in 2014, using APIs from WWT (Microsoft Research, now AAS) and CDS (Aladin)

Images-on-the-Sky relies on the astrometry.net, Zooniverse, IOP/AAS Astronomy Image Explorer and WorldWide Telescope platforms, and it is funded by the **American Astronomical Society**, in addition to the NASA ADAP grant.

These projects rely on open source sofware, primarily hosted on **GitHub**.

PI to contact for more information
Alyssa Goodman, Harvard
agoodman@cfa.harvard.edu

SEAMLESS ASTRONOMY
Linking scientific data, publications, and communities

your handout