Exploring High-Dimensional Data in Astronomy, Genomics, and beyond, using glue

Alyssa Goodman
Robert Wheeler Willson Professor, Astronomy, Harvard University
& President, glue solutions, inc.

Jonathan Foster
Chief Technology Officer, glue solutions, inc.

AstroMed

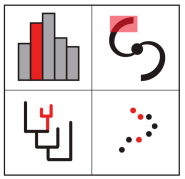
AMERICAN ASTRONOMICAL SOCIETY WORLDWIDE TELESCOPE

glue
multidimensional data exploration

glupyter
multidimensional data exploration

glue genes

NASA NSF HARVARD GORDON AND BETTY MOORE FOUNDATION The Jackson Laboratory glue solutions inc.

What is  glue?

multidimensional data exploration

It’s not an acronym.

It is open-source software that

glues data,

glues graphs &

glues tools.

data

numbers (tables, arrays, spreadsheets)

images & maps (FITS, JPEG, GIS and more)

data cubes (3D, 4D, and more)

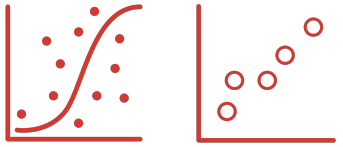
data files' common attributes are glued

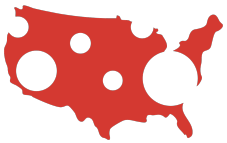
avoiding the need to merge data files

“graphs”

common statistical graphics

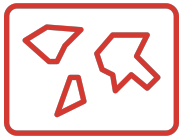
(scatterplots, histograms, tables, curves, overlays)





maps & images

(greyscale, color, contours, layer control…)

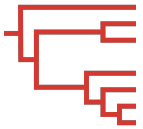


3D displays

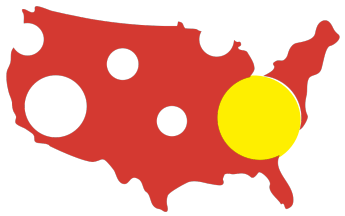
(scatter plots, volumetric rendering, sliders…)

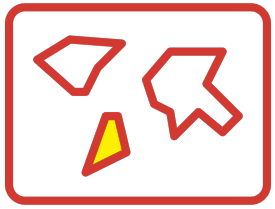
specialized & custom charts

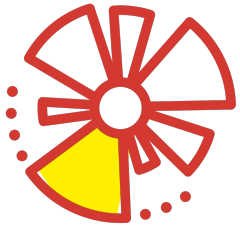
(dendrograms, polar plots, + domain-specific options)



selections propagate across all graphs



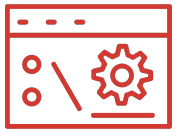




for real-time data exploration & insight

tools

plug-ins (user-defined formats, plots, layouts…)



web services (across domains)

command-line (built-in terminal, scriptable)

for easy customization

glues data,
glues graphs &
glues tools.

glueviz.org

BONUS: **save**, **share**, or **publish** what you learn—

save “sessions” to continue where you left off

export graphics

use/export to Jupyter environments

export to plot.ly (javascript)

export to augmented reality

learn how at glueviz.org.

glueviz.org

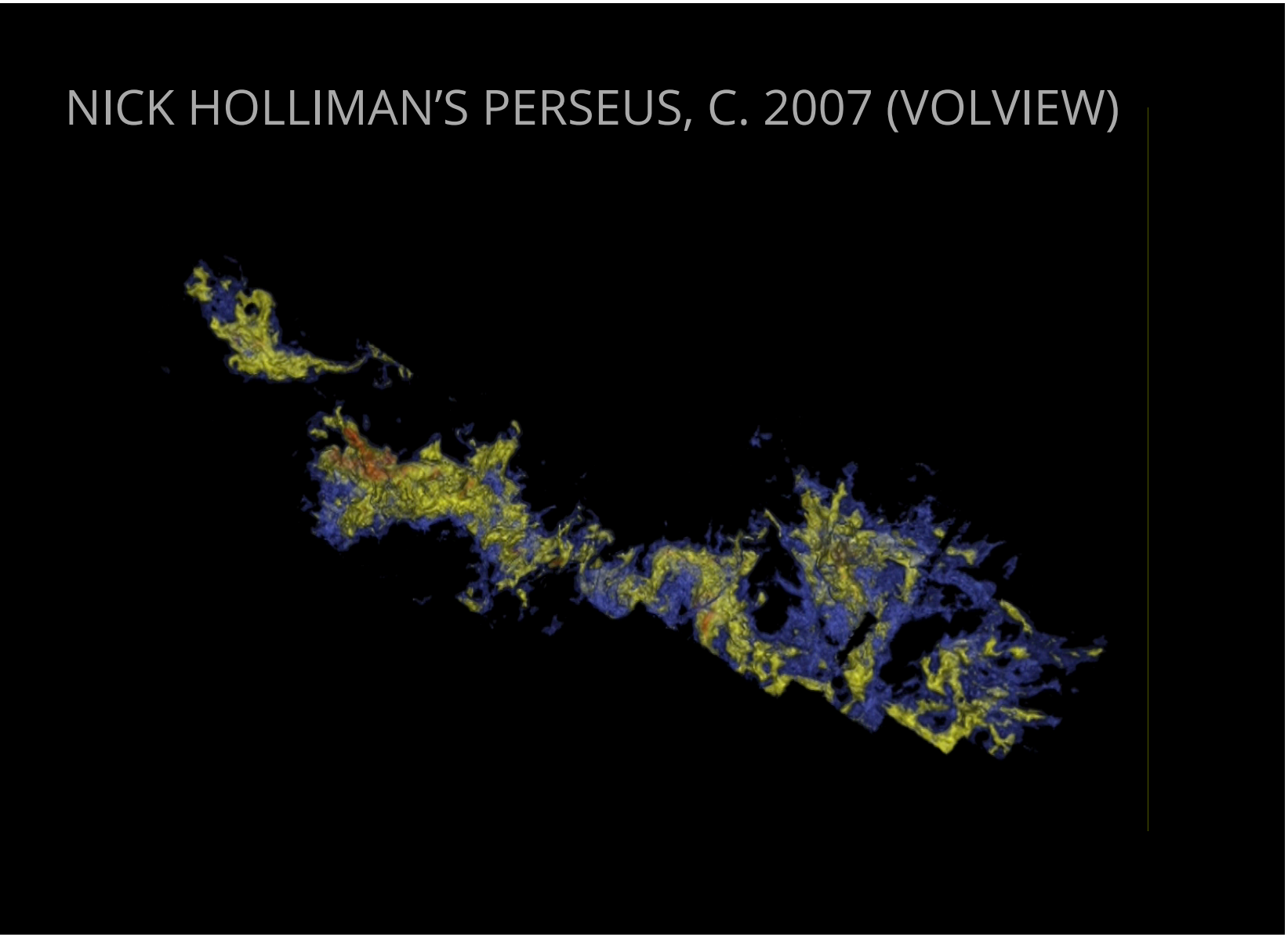
supported by

glue solutions inc.

GORDON AND BETTY MOORE FOUNDATION

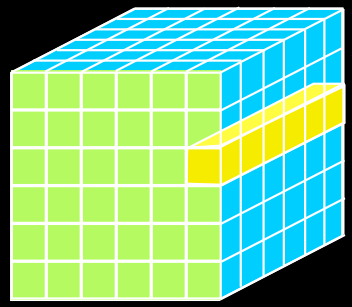
Exploring High-Dimensional Data in Astronomy, Genomics, and beyond, using glue

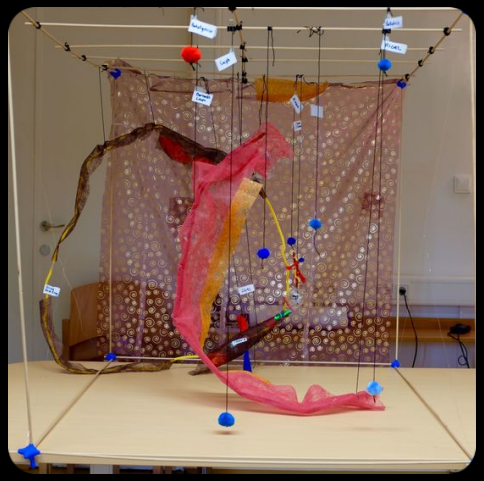
NICK HOLLIMAN'S PERSEUS, C. 2007 (VOLVIEW)

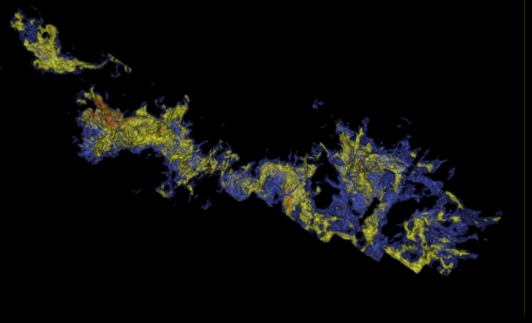


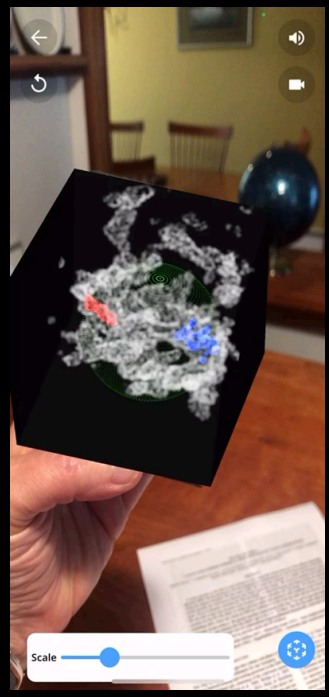
AstroMed

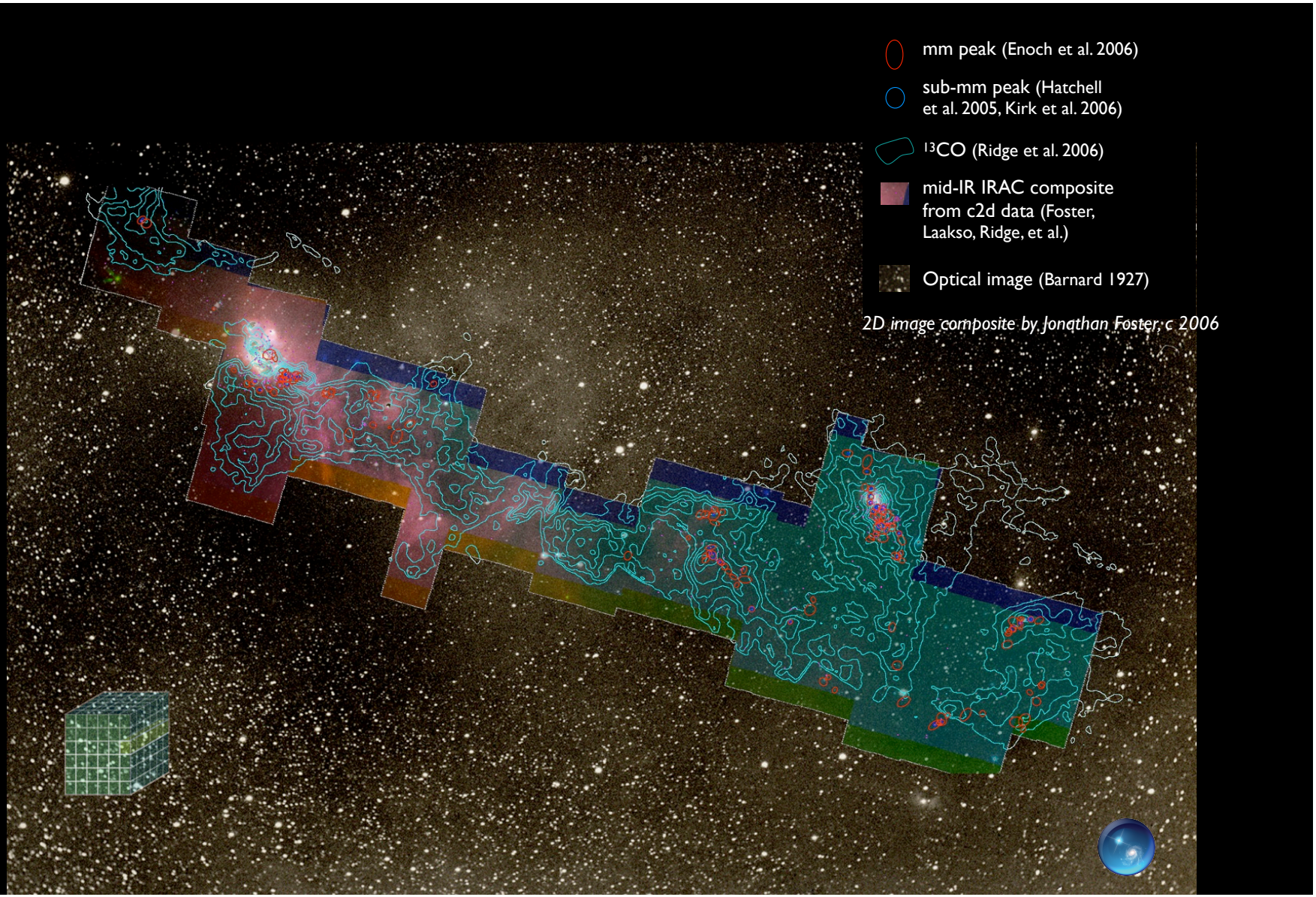
“DATA, DIMENSIONS, DISPLAY”





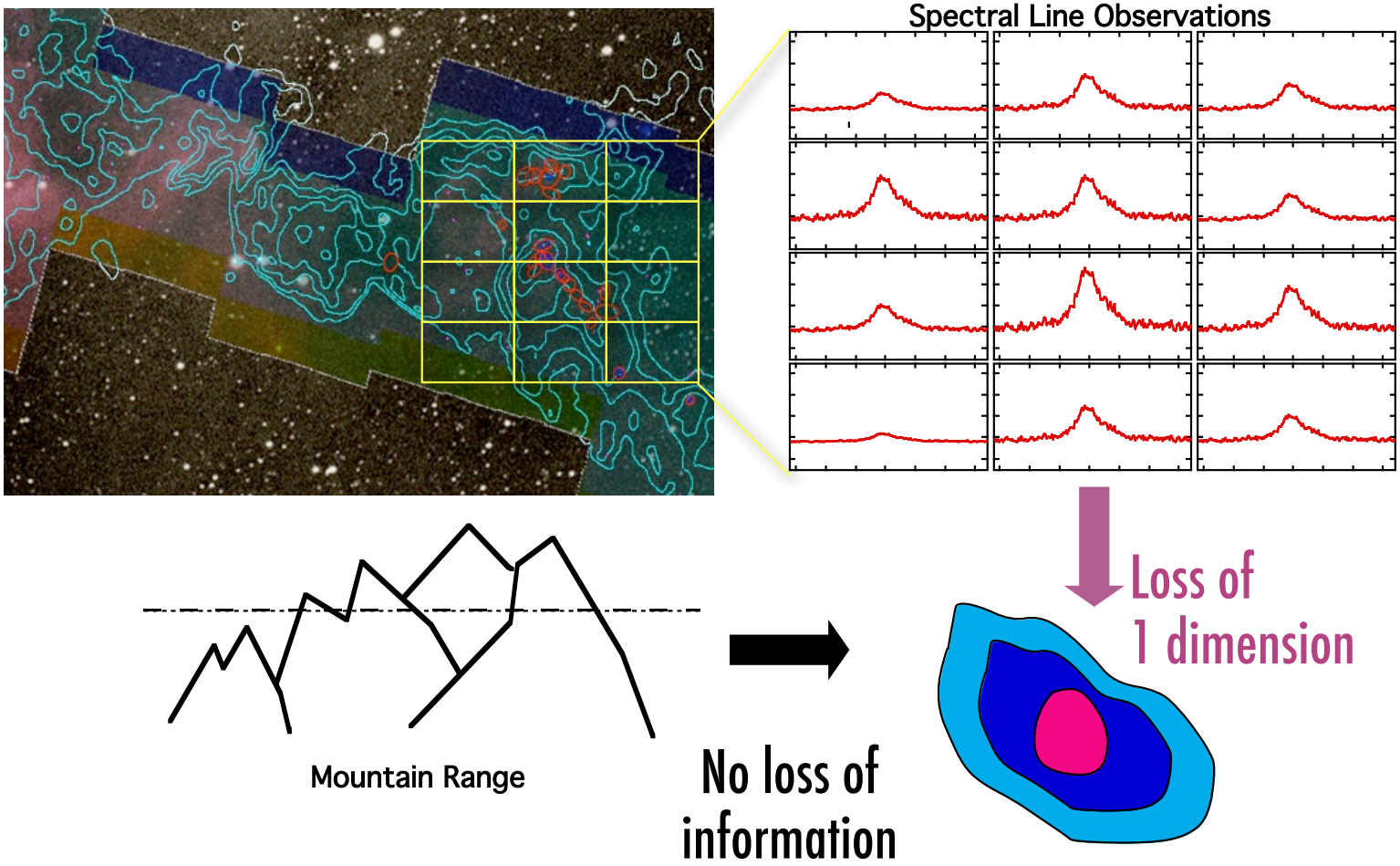





mm peak (Enoch et al. 2006)
sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)
^{13}CO (Ridge et al. 2006)
mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al.)
Optical image (Barnard 1927)
2D image composite by Jonathan Foster, c 2006

AstroMed

Spectral-line mapping (what's "*p-p-v*" space?)



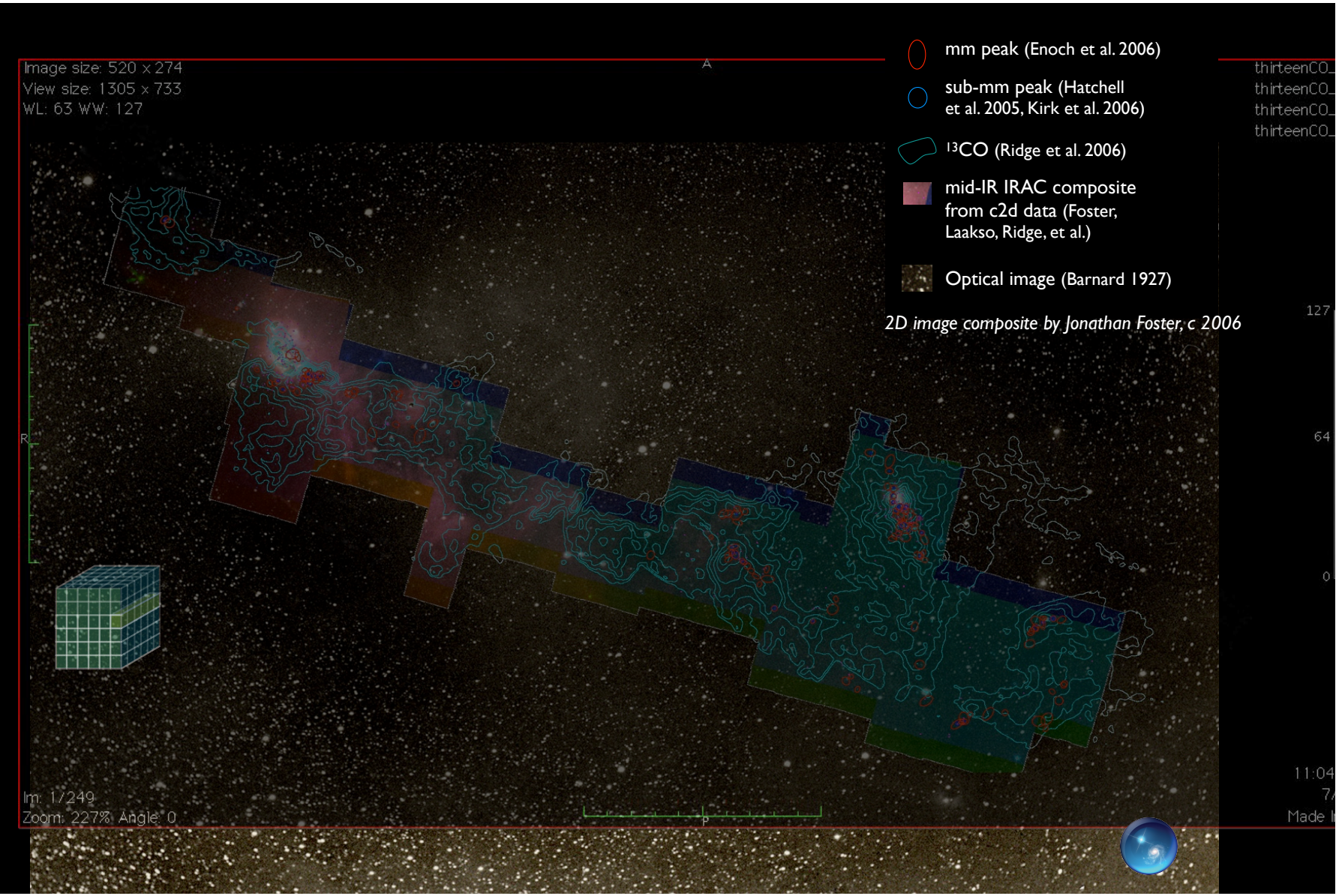
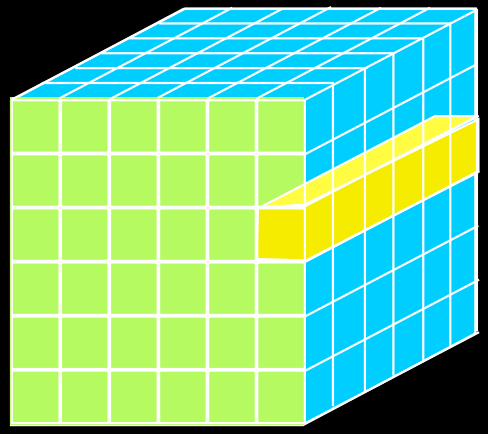


Image size: 520 x 274
View size: 1305 x 733
WL: 63 WW: 127
mm peak (Enoch et al. 2006)
sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)
13CO (Ridge et al. 2006)
mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al.)
Optical image (Barnard 1927)
2D image composite by Jonathan Foster, c 2006
thirteenCO_
thirteenCO_
thirteenCO_
thirteenCO_
127
64
0
11:04
Im: 1/249
Zoom: 227% Angle: 0

AstroMed



DATA-DIMENSIONS-DISPLAY

1D: Columns = “Spectra”, “SEDs” or “Time Series” (x-y Graphs)
2D: Faces or Slices = “Images”
3D: Volumes = “3D Renderings”, “2D Movies”
4D: Time Series of Volumes = “3D Movies”

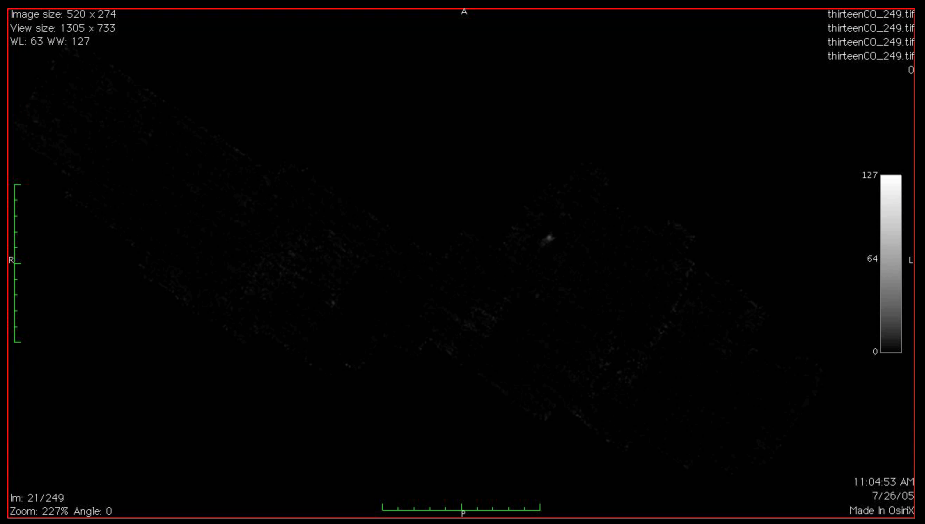
ASTRONOMICAL MEDICINE

AstroMed

"Keith"

"z" is depth into head

"Perseus"



"z" is line-of-sight velocity

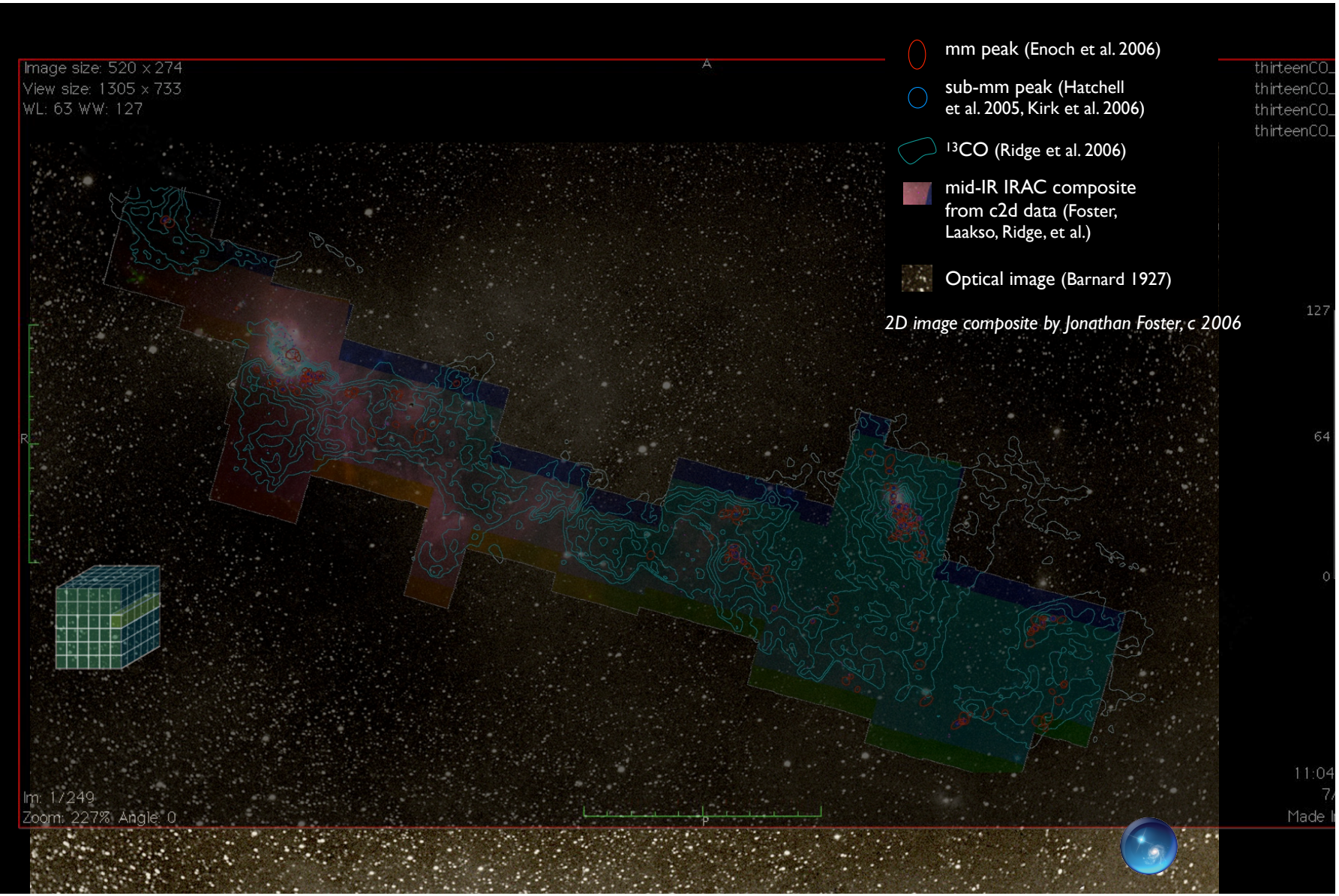
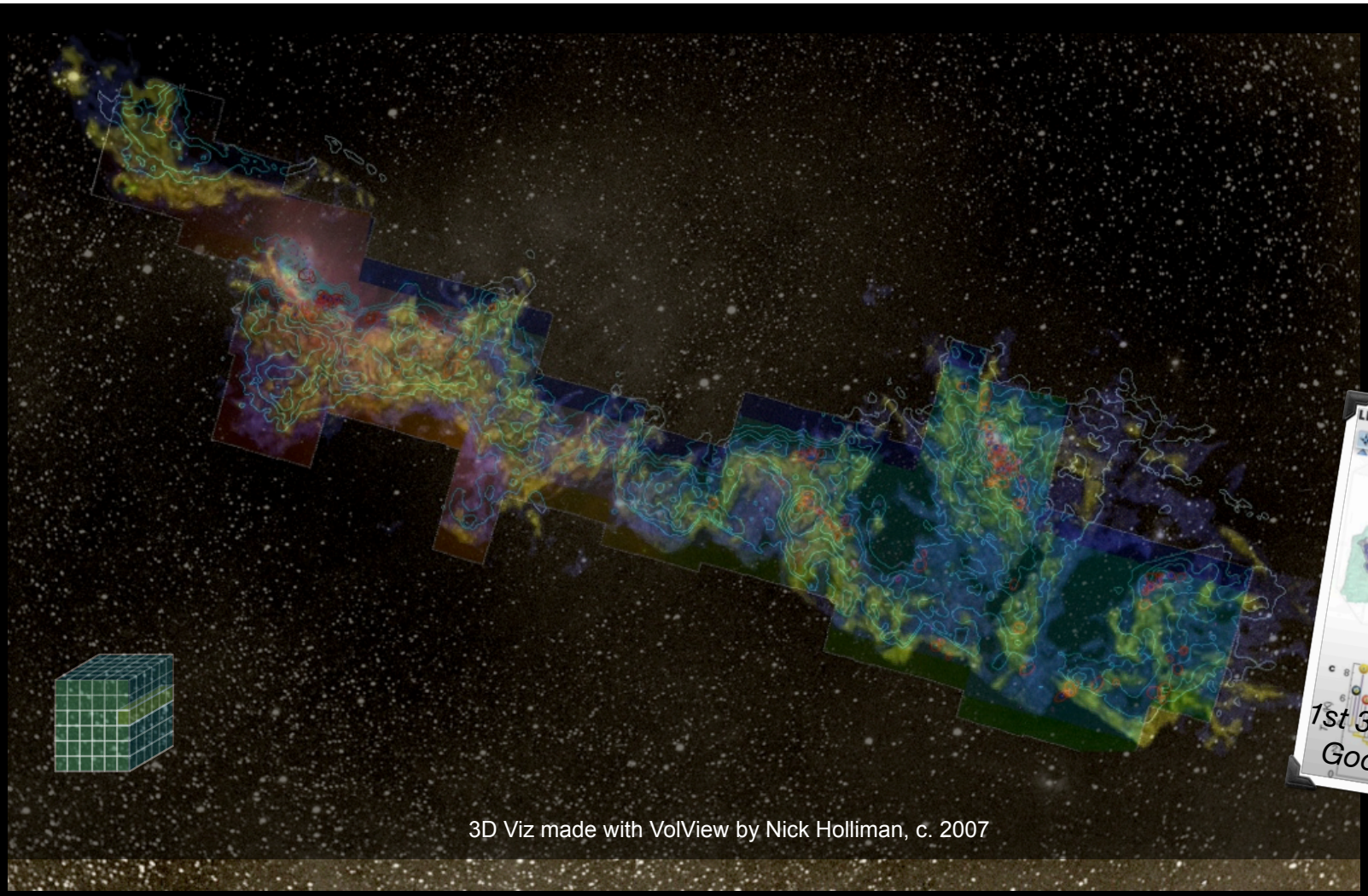
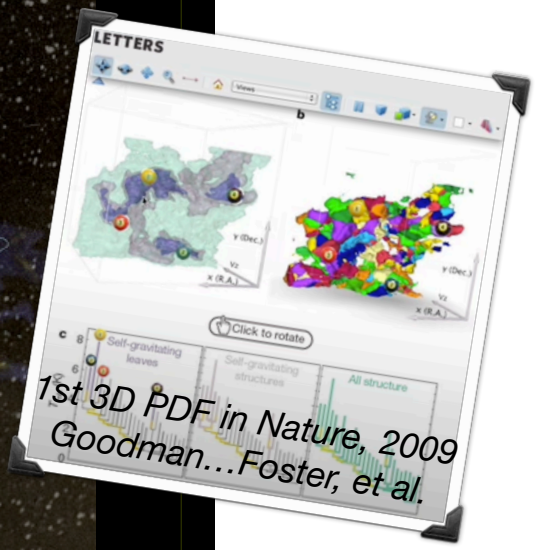


Image size: 520 x 274
View size: 1305 x 733
WL: 63 WW: 127
mm peak (Enoch et al. 2006)
sub-mm peak (Hatchell et al. 2005, Kirk et al. 2006)
13CO (Ridge et al. 2006)
mid-IR IRAC composite from c2d data (Foster, Laakso, Ridge, et al.)
Optical image (Barnard 1927)
2D image composite by Jonathan Foster, c 2006
thirteenCO_
thirteenCO_
thirteenCO_
thirteenCO_
127
64
0
11:04
Im: 1/249
Zoom: 227% Angle: 0

AstroMed



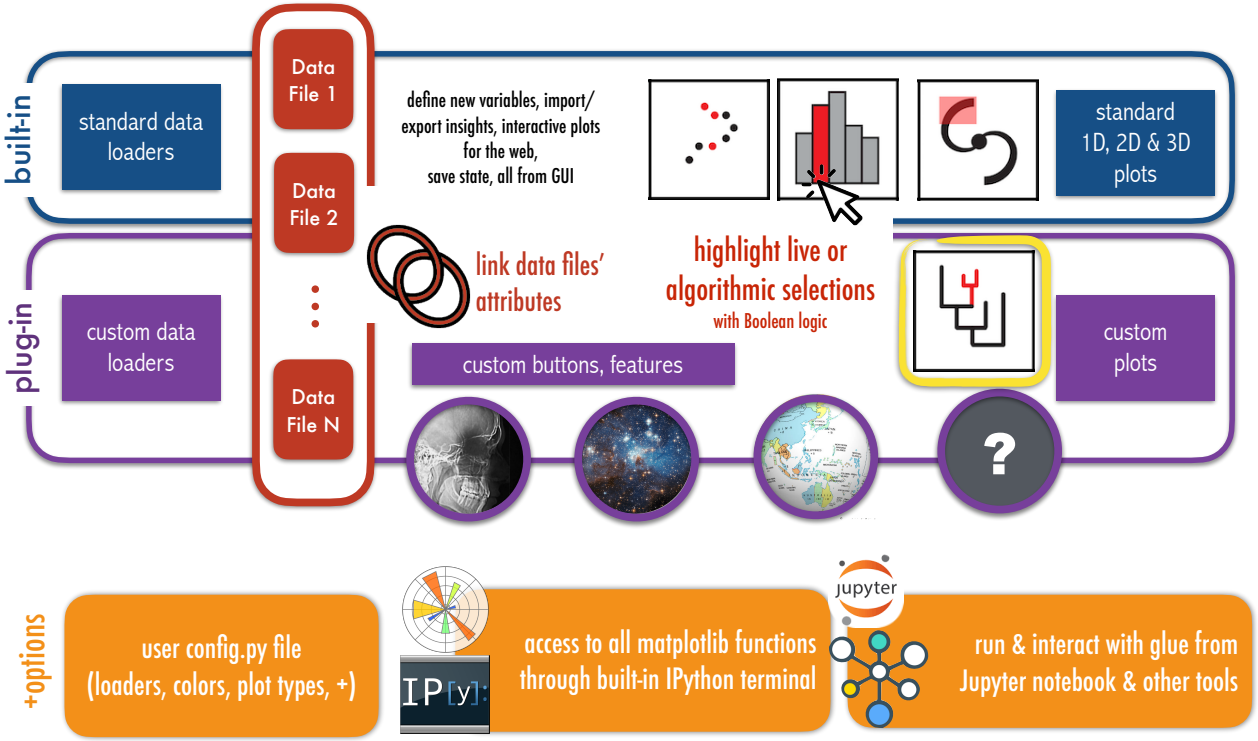
3D Viz made with VolView by Nick Holliman, c. 2007

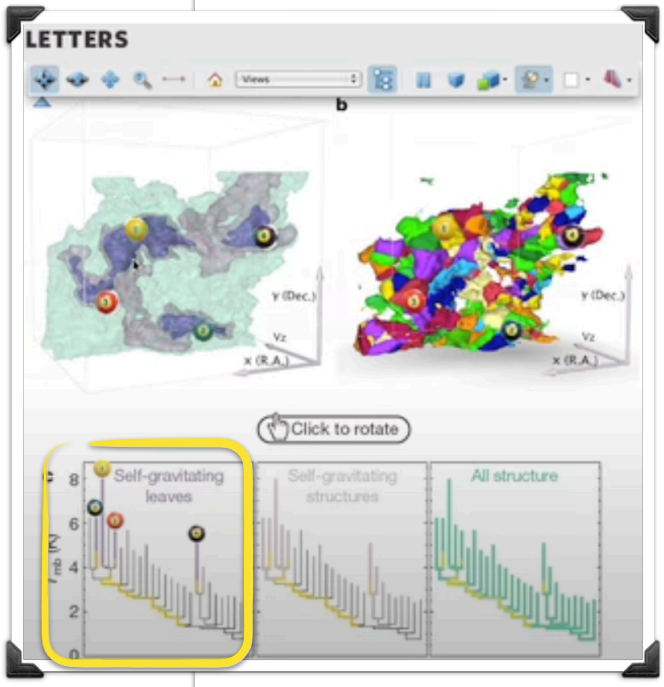


1st 3D PDF in Nature, 2009
Goodman…Foster, et al.

AstroMed

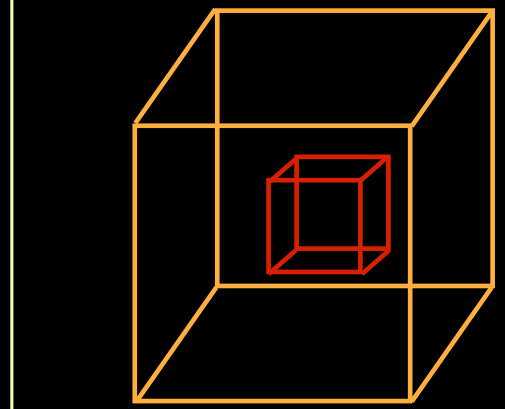
COMPLETE


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
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

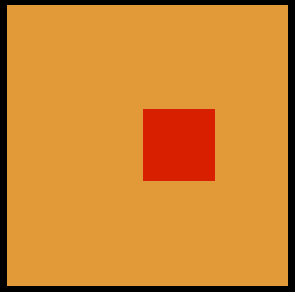

LETTERS
Click to rotate
Self-gravitating leaves
Self-gravitating structures
All structure

glueviz.org

LINKED VIEWS OF HIGH-DIMENSIONAL DATA



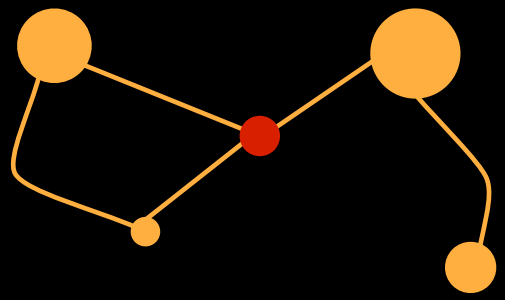
3D



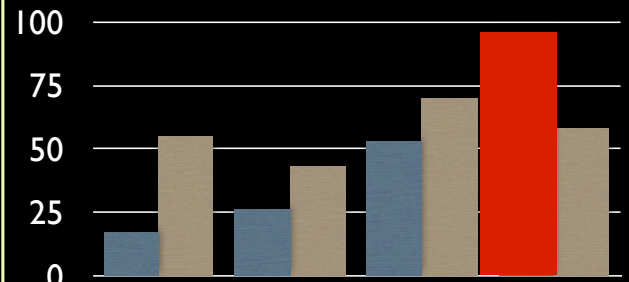
2D

John Tukey

Data Abstraction

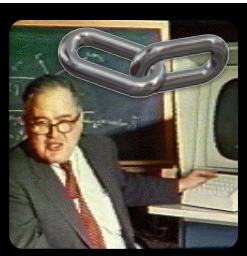


Statistics



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

JOHN TUKEY'S LEGACY



PRIM-9

PRIM-H

DataDesk

glue

XGobi → GGobi

RGGobi

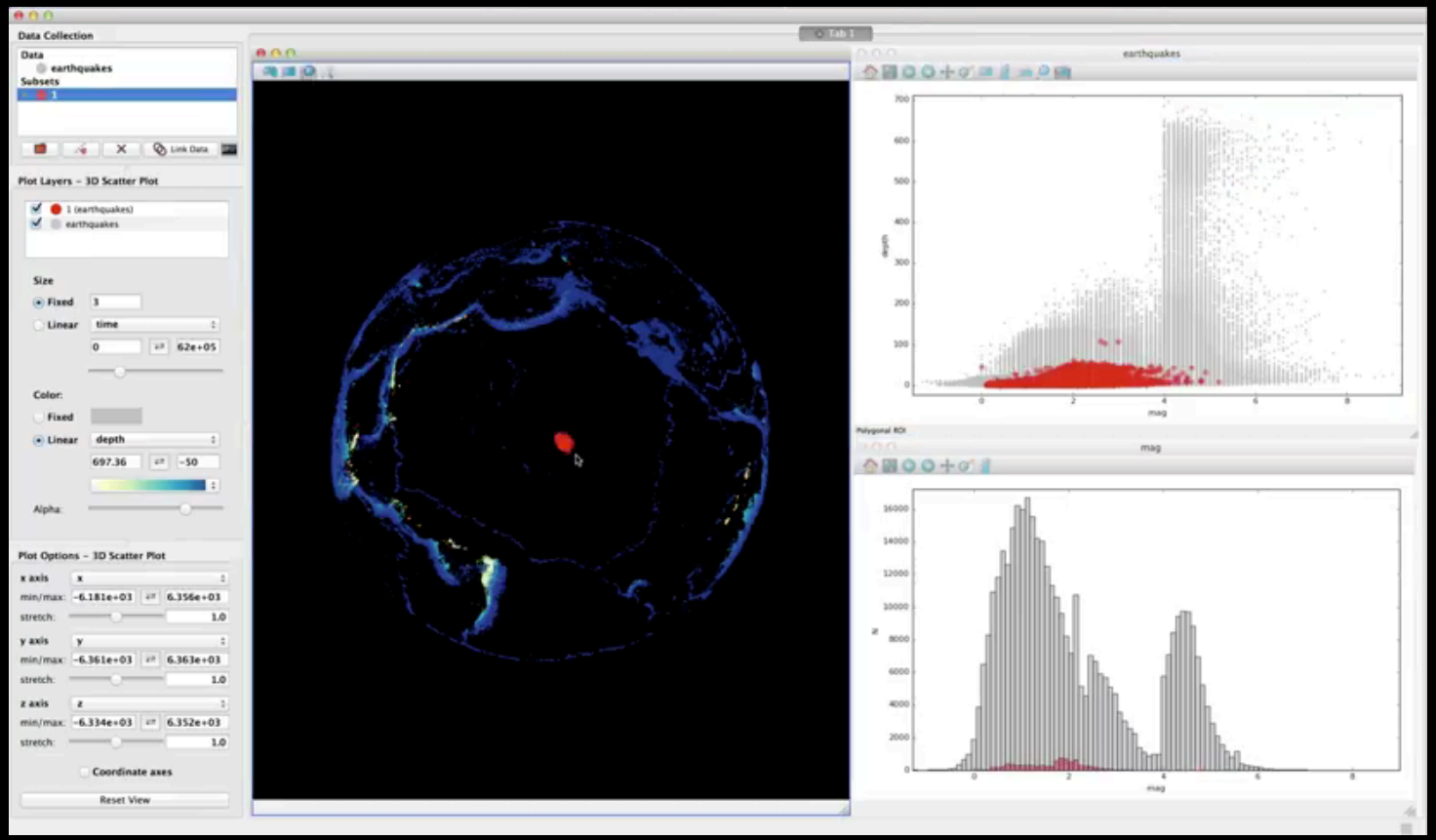
Microsoft Power BI

Spotfire

Polaris → tableau SOFTWARE

1970 1980 1990 2000 2010

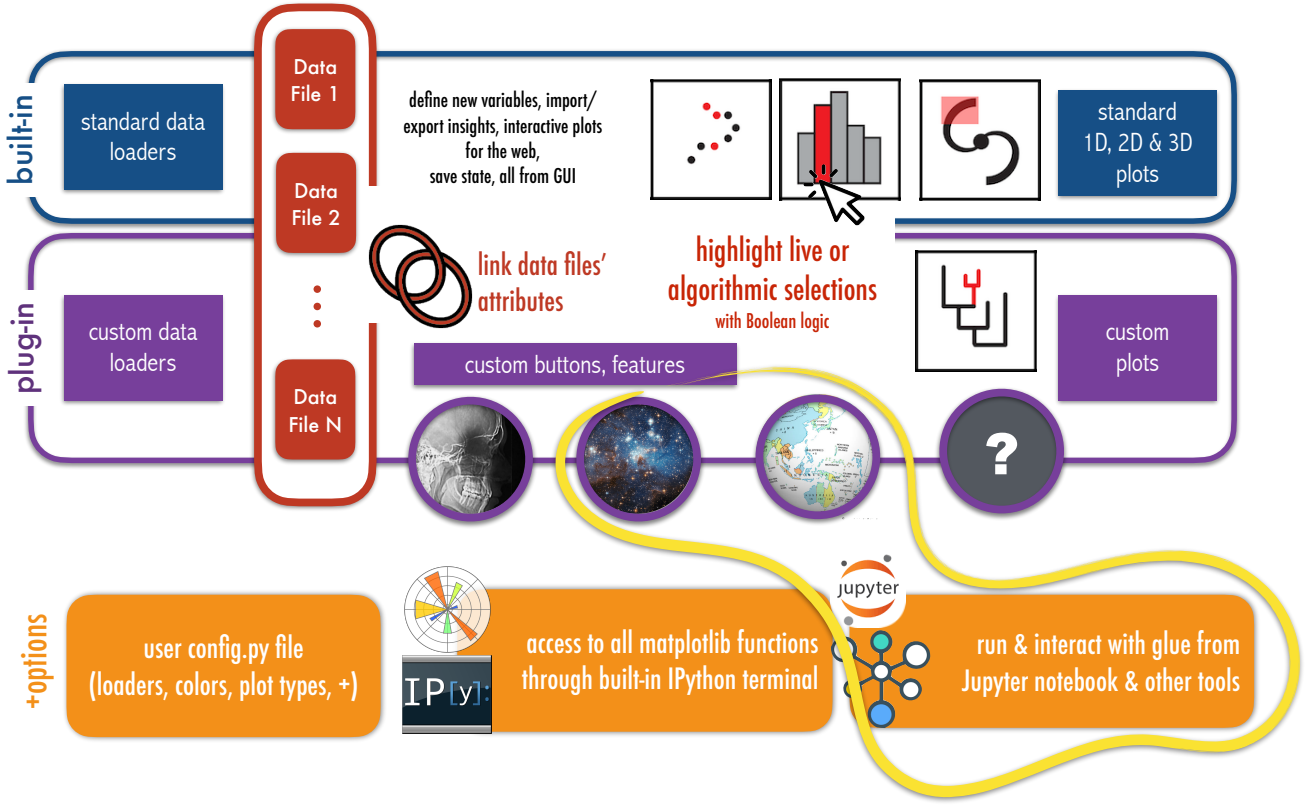
LINKED VIEWS OF HIGH-DIMENSIONAL DATA (IN PYTHON)



glue
multidimensional data exploration

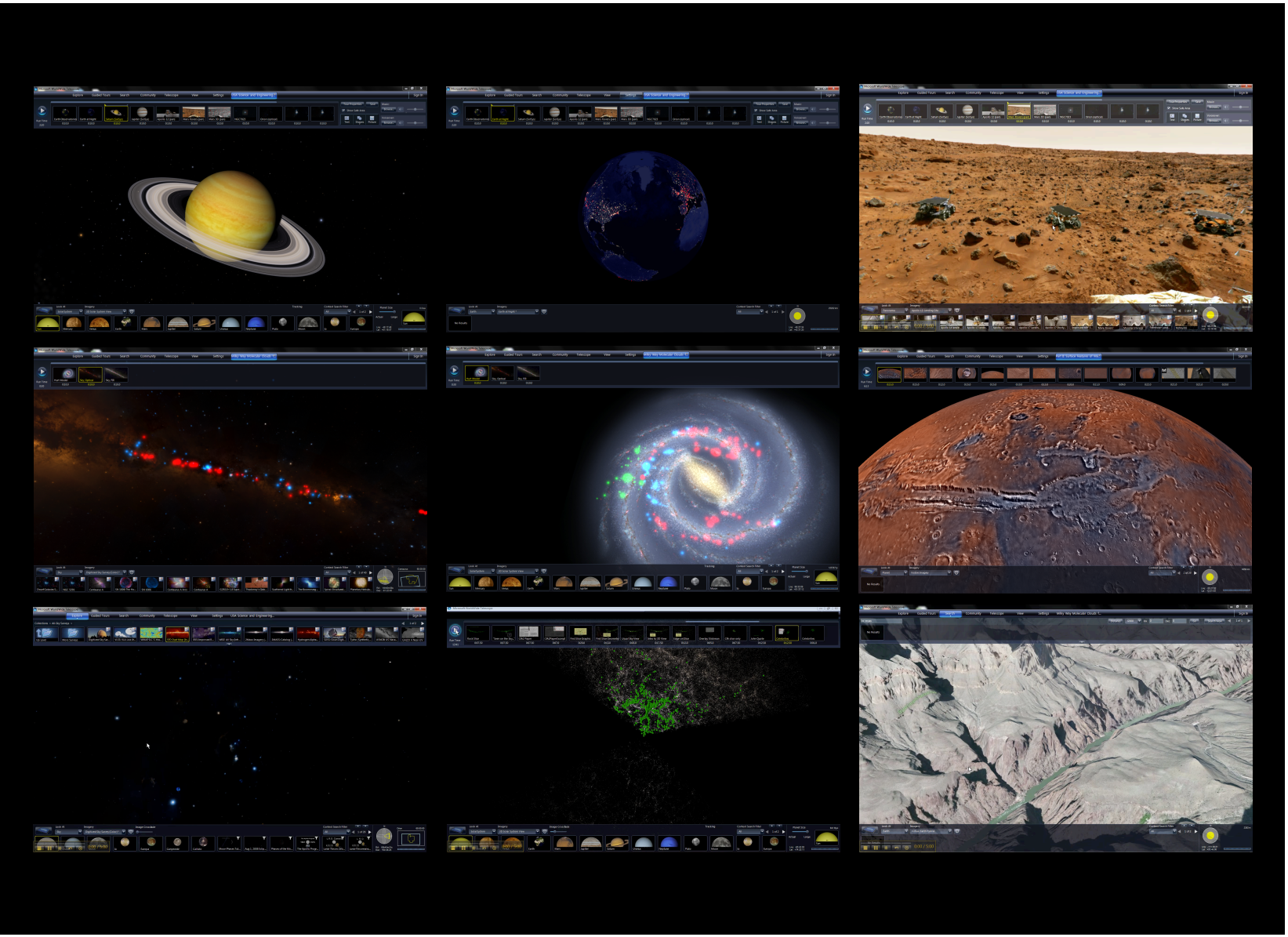
video by Tom Robitaille, lead glue developer

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



glue
multidimensional data exploration

glueviz.org



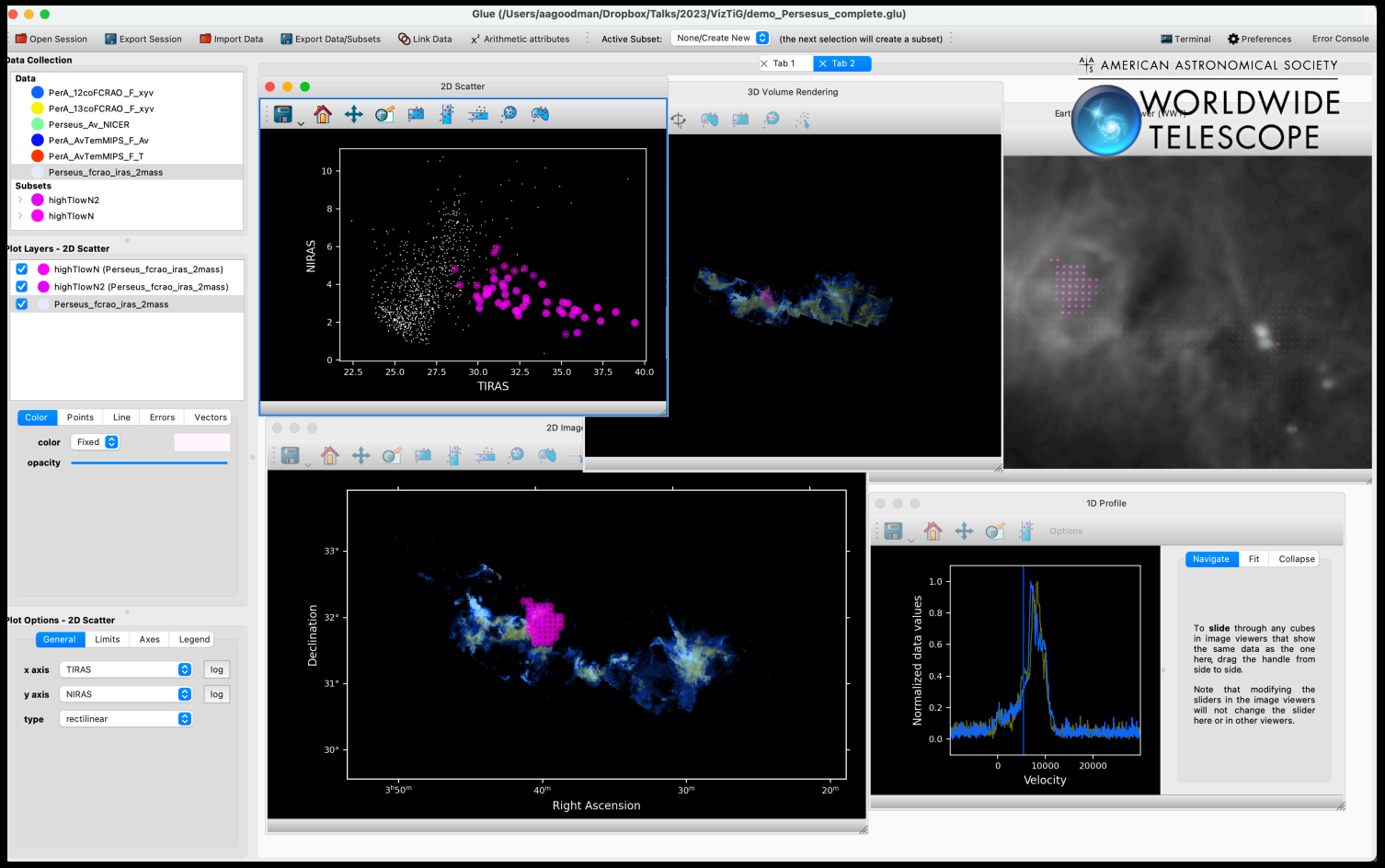
AMERICAN ASTRONOMICAL SOCIETY
WORLDWIDE TELESCOPE

worldwidetelescope.org

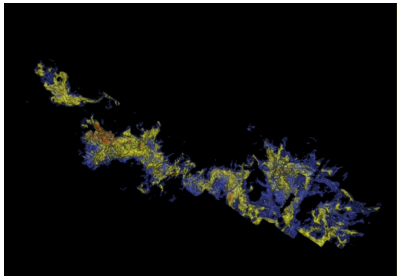
AMERICAN ASTRONOMICAL SOCIETY

WORLDWIDE TELESCOPE

worldwidetelescope.org



[demo; a 1-year data synthesis study in 5 minutes=Ridge et al. 2006]



glue
multidimensional data exploration

The Astrophysical Journal, 643:932–944, 2006 June 1

THE COMPLETE NATURE OF THE WARM DUST SHELL IN PERSEUS

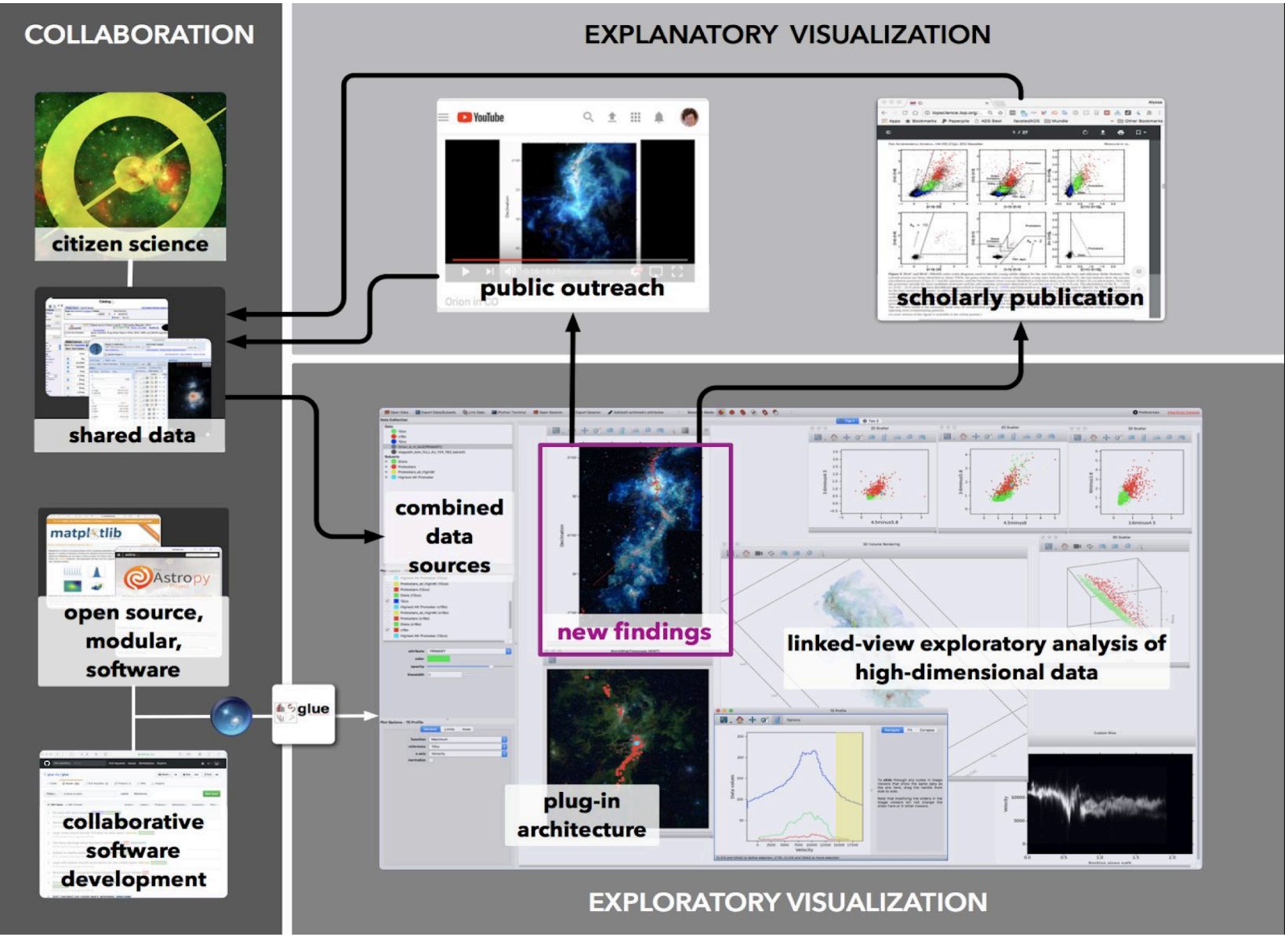
Naomi A. Ridge, Scott L. Schnee, Alyssa A. Goodman, and Jonathan B. Foster
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138; nridge@cfa.harvard.edu, sschnee@cfa.harvard.edu, agoodman@cfa.harvard.edu
Received 2005 April 28; accepted 2006 January 19

ABSTRACT

The Perseus molecular cloud complex is a ≥30 pc long chain of molecular clouds most well known for the two star-forming clusters NGC 1333 and IC 348 and the well-studied outflow source in B5. However, when studied at mid- to far-infrared wavelengths, the region is dominated by a ~10 pc diameter shell of warm dust, likely generated by an H II region caused by the early-B star HD 278942. Using a revised calibration technique the COMPLETE team has produced high-sensitivity temperature and column density maps of the Perseus region from *IRAS* Sky Survey Atlas (ISSA) 60 and 100 μm data. In this paper, we combine the ISSA-based dust-emission maps with other observations collected as part of the COMPLETE Survey, along with archival Hα and *MSX* observations. Molecular line observations from FCRAO and extinction maps constructed by applying the NICER method to the 2MASS catalog provide independent estimates of the "true" column density of the shell. Hα emission in the region of the shell confirms that it is most likely an H II region located behind the cloud complex, and 8 μm data from *MSX* indicate that the shell may be interacting with the cloud. Finally, the two polarization components seen toward background stars in the region by Goodman et al. can be explained by the association of the stronger component with the shell. If confirmed, this would be the first observation of a parsec-scale swept-up magnetic field.

Subject headings: dust, extinction — H II regions — infrared: ISM — ISM: individual (G159.6-18.5) — radio lines: ISM

Online material: color figure

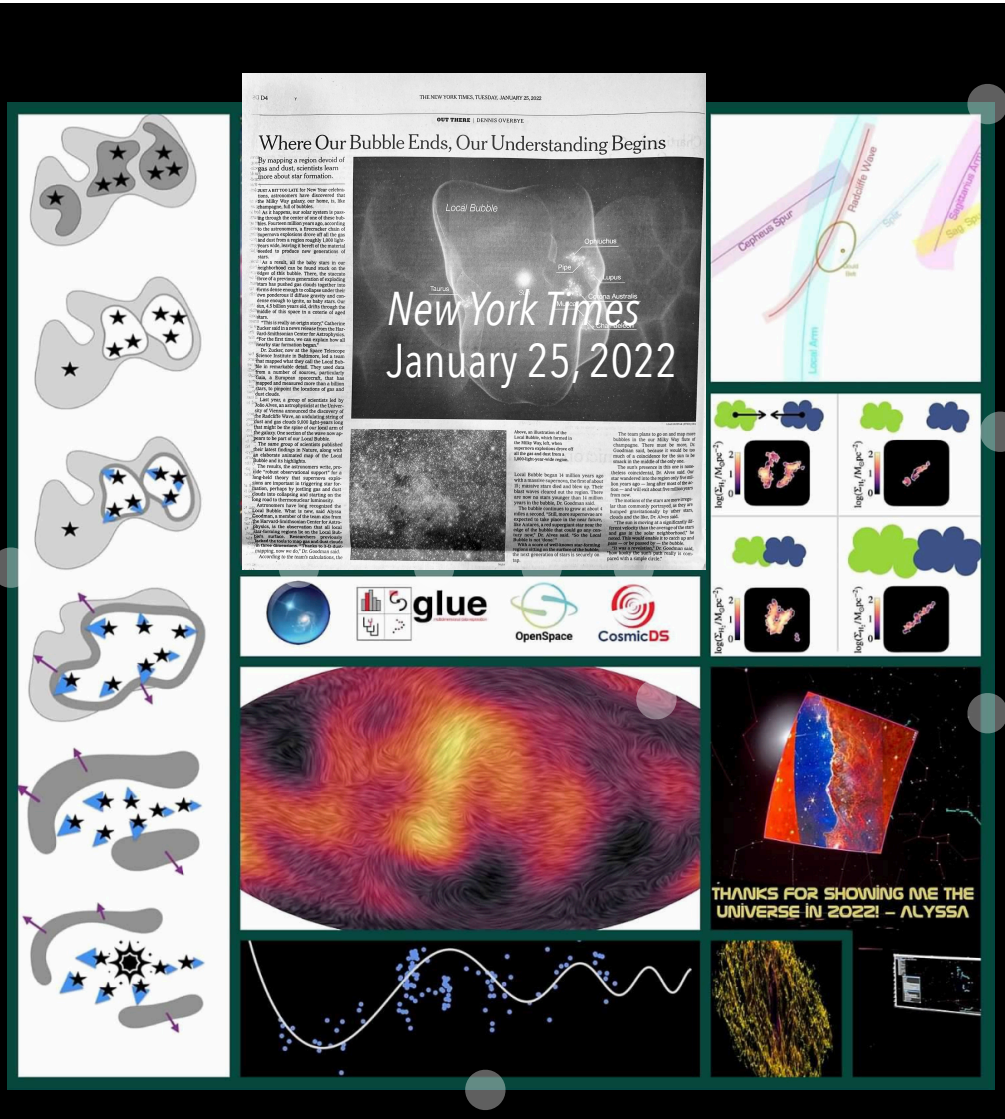


glue
multidimensional data exploration

New Thinking on, and with, Data Visualization

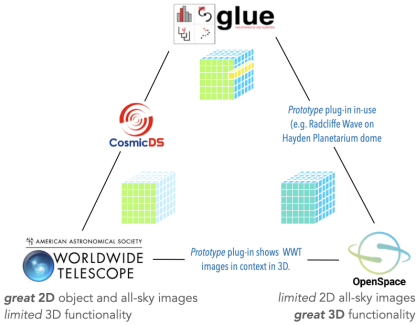
Alyssa A. Goodman, Harvard University
Michelle A. Borkin, Northeastern University
Thomas P. Robitaille, Aperio Software Ltd.

arxiv.org/abs/1805.11300



glue
multidimensional data exploration

great **1D**, **2D** and **3D** data manipulation, flexible architecture facilitating plug-ins, data transfer, and interactive data exploration; "glupyter" flavor runs in web pages



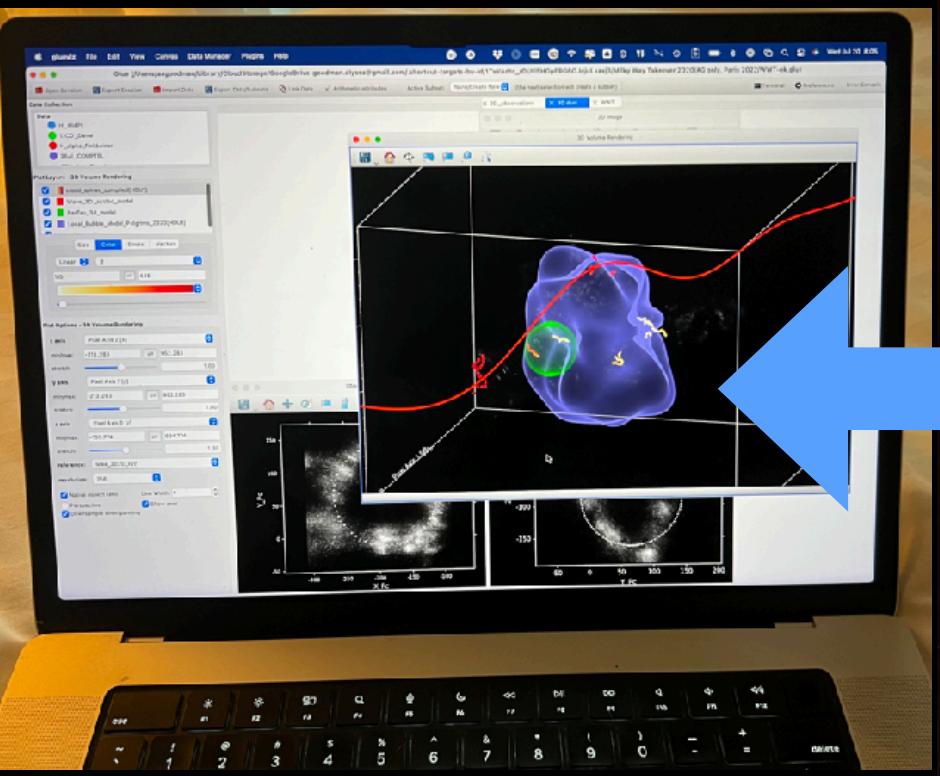
great **2D** object and all-sky images
limited 3D functionality

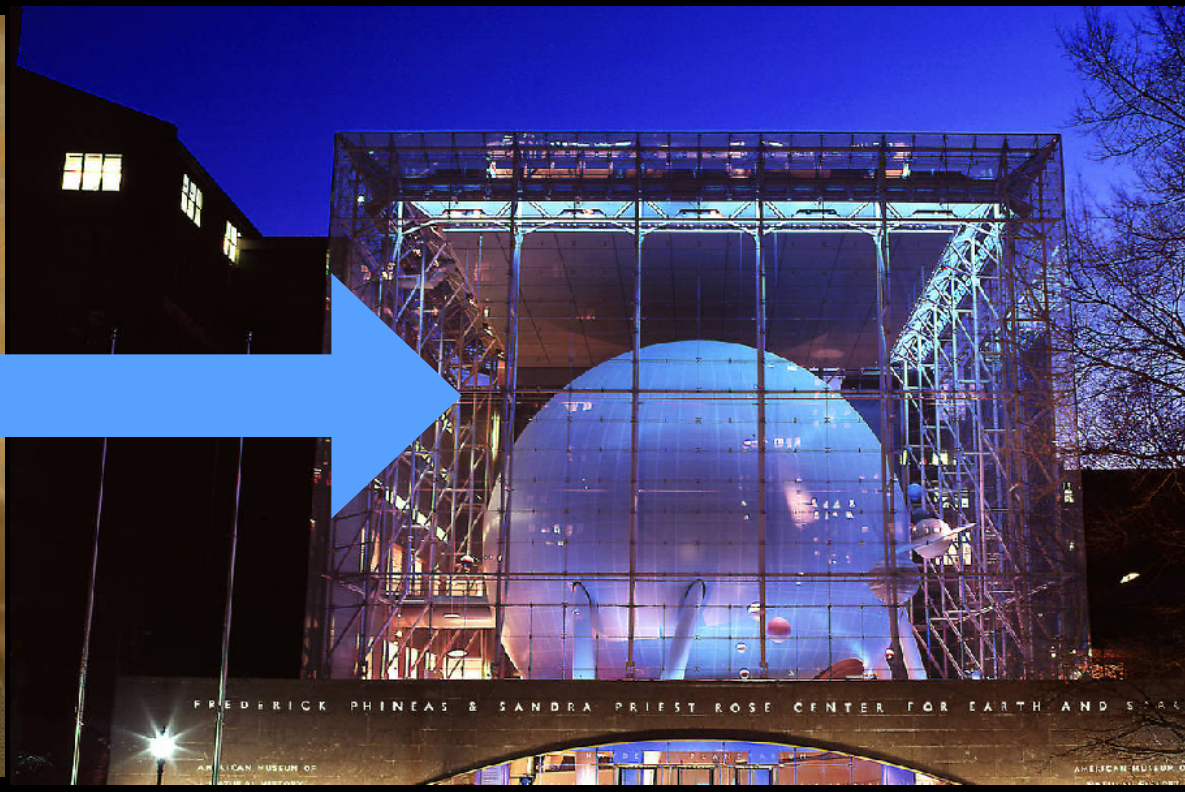
limited 2D all-sky images
great **3D** functionality

glue

OpenSpace

AMERICAN ASTRONOMICAL SOCIETY
WORLDWIDE TELESCOPE





NASA

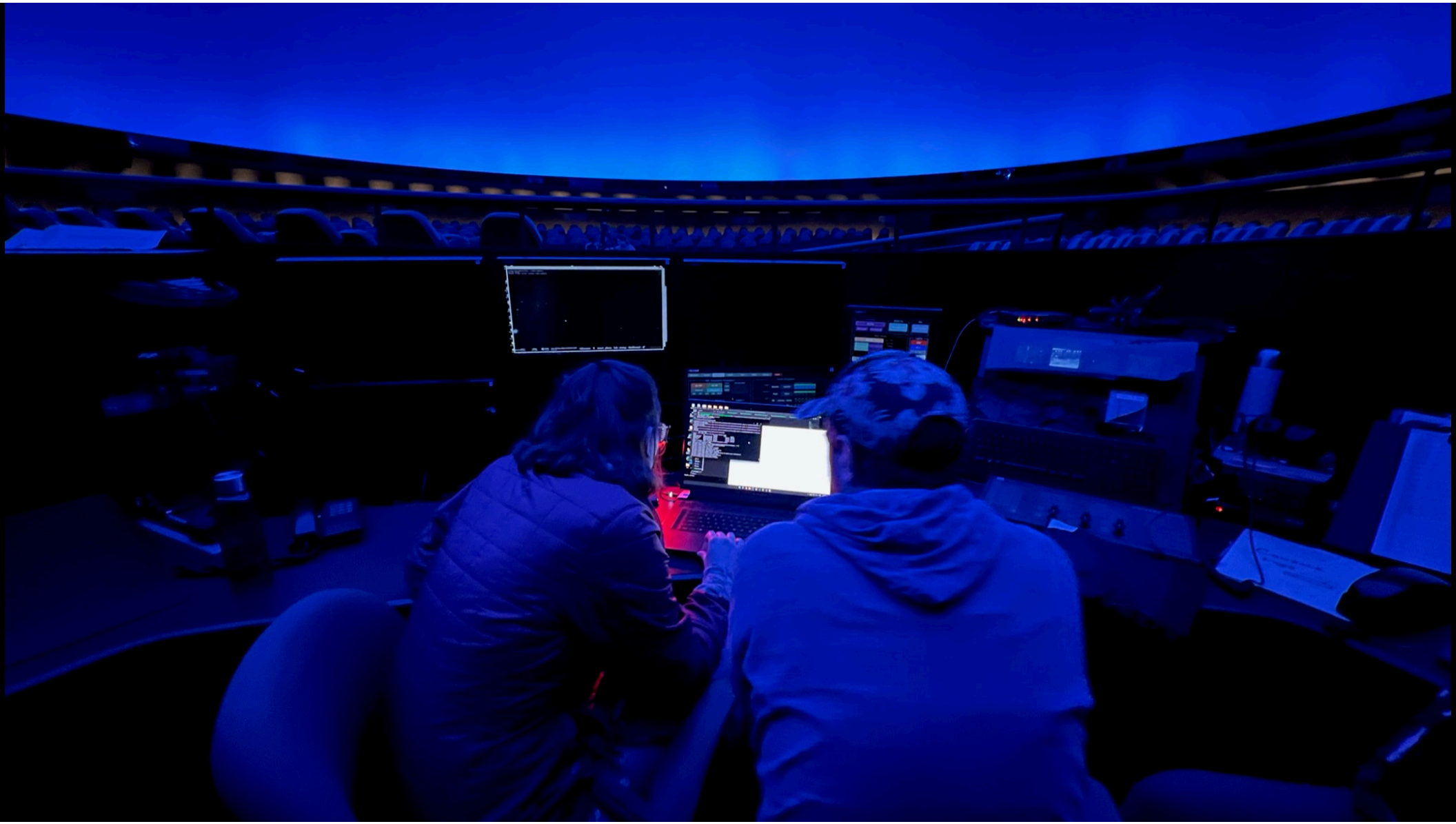
NSF

Welcome to a new view of the Milky Way... in 3D!

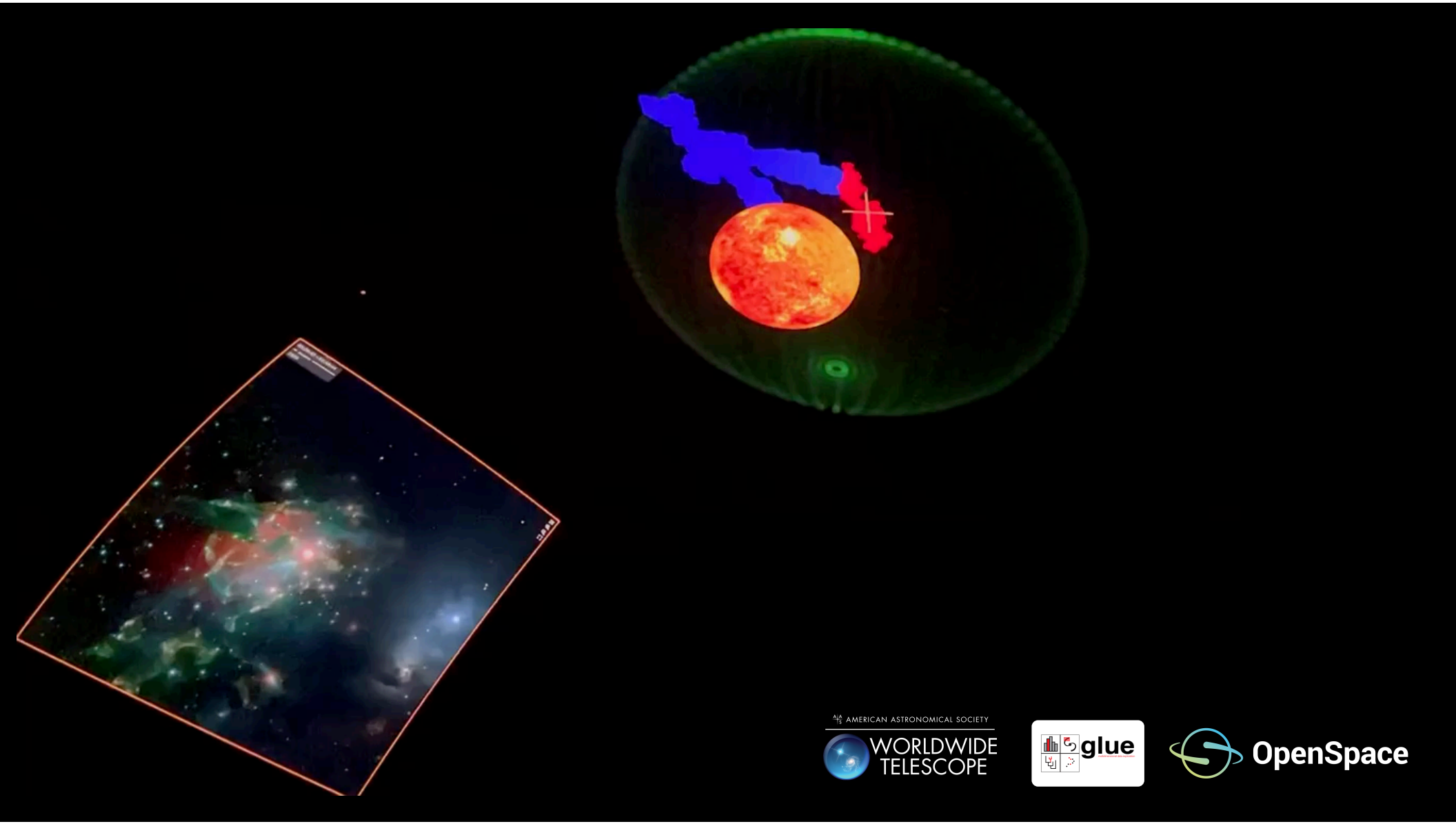
Soon, milkyway3d.org will serve as a hub for the interconnected set of outreach, education, and research resources that will result from the interconnections we're in the process of making.

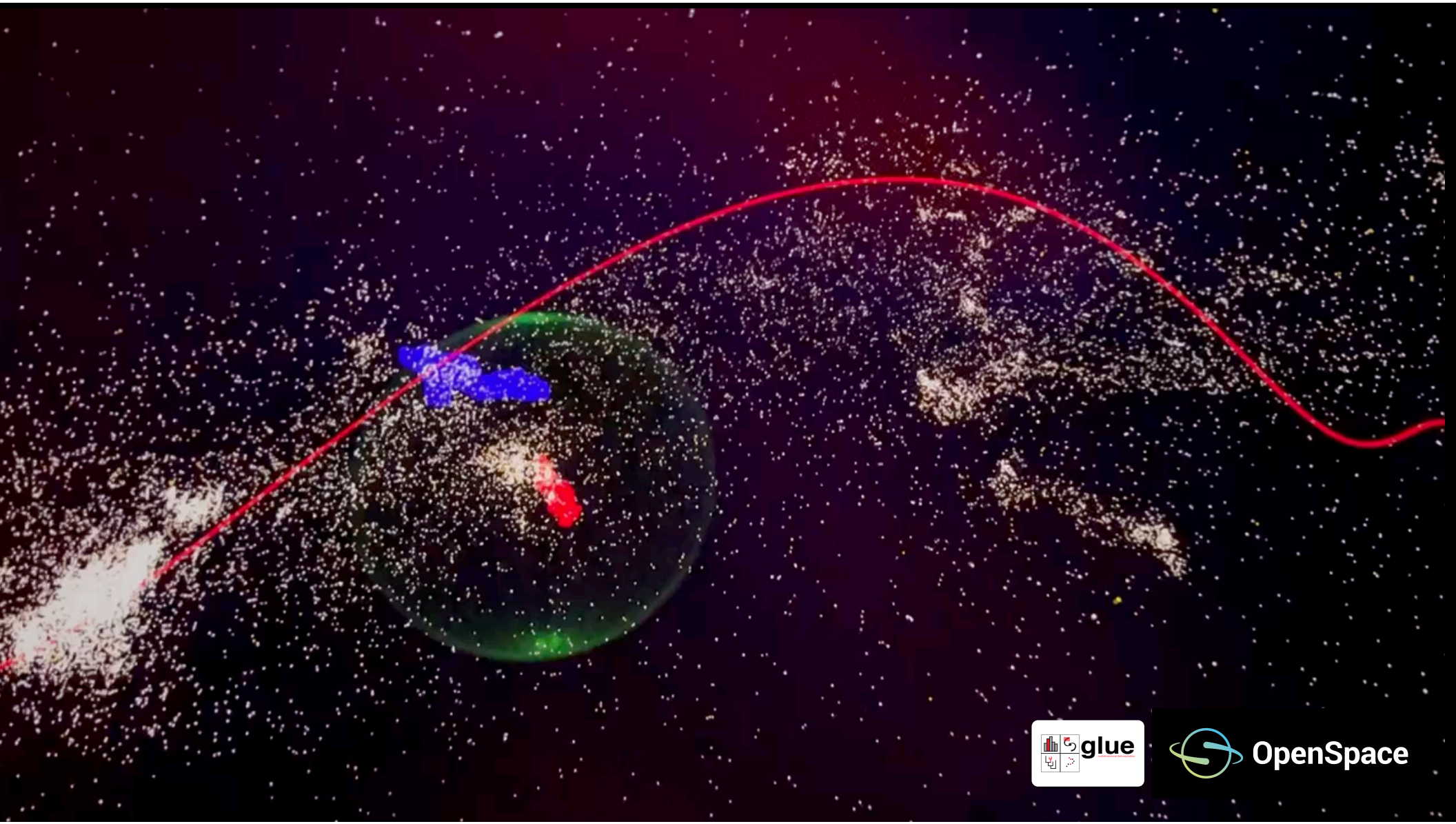
Our project includes new software development; approaches to data sharing; and scientific research questions propelling our collaboration forward.

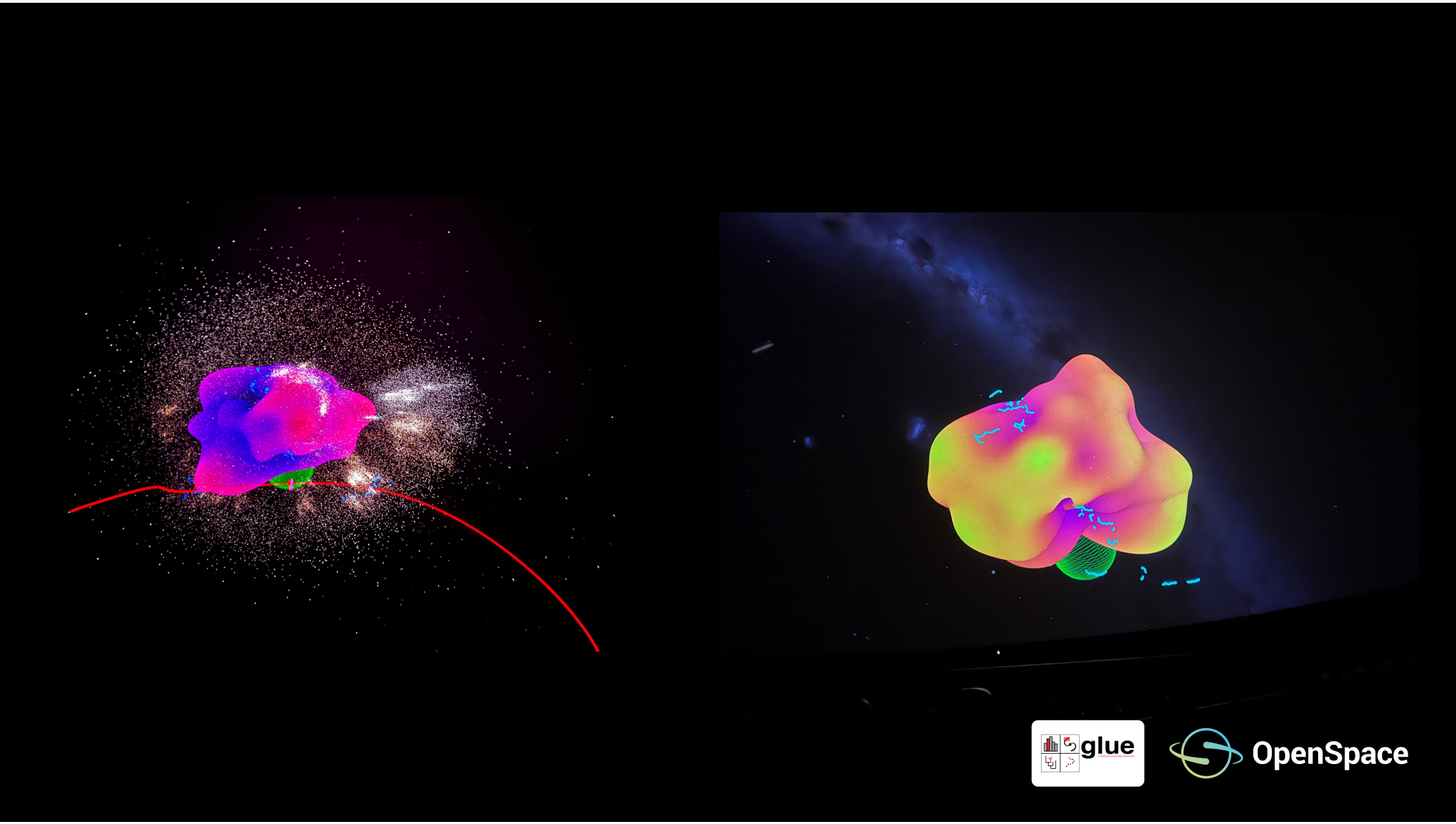
milkyway3D.org



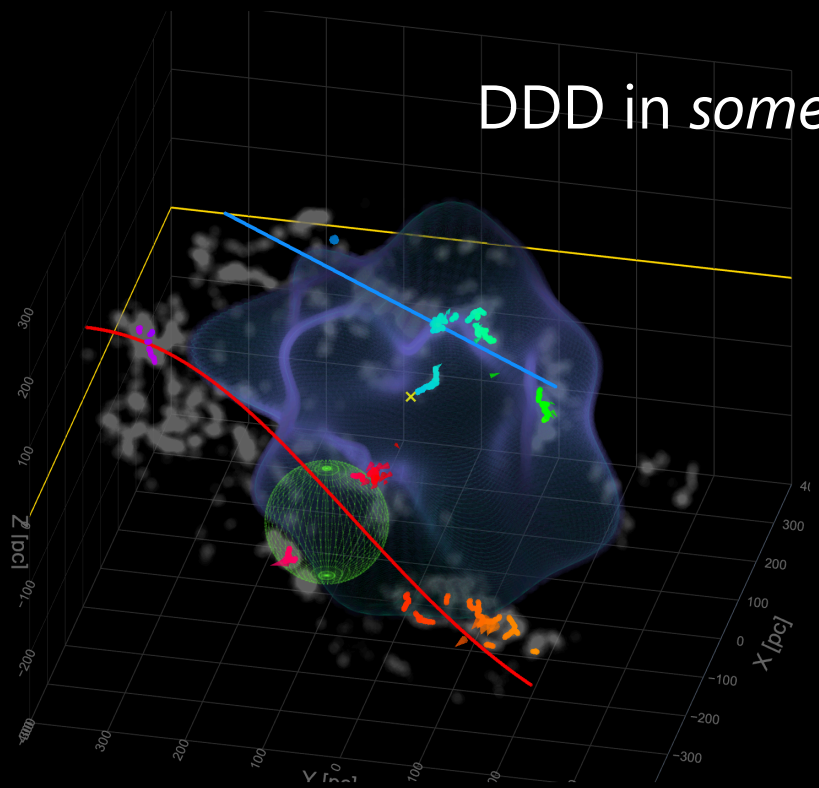

glue
OpenSpace


AMERICAN ASTRONOMICAL SOCIETY
WORLDWIDE TELESCOPE
glue
OpenSpace

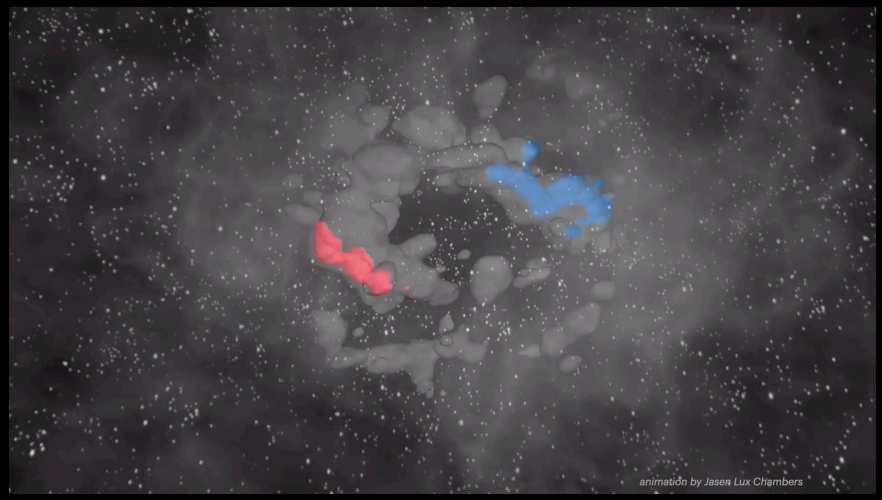

glue
OpenSpace

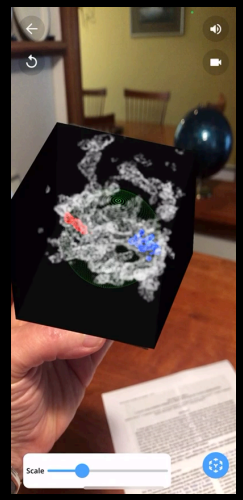

glue
OpenSpace

DDD in *some* of the Local ISM, recently



Zucker et al. 2022

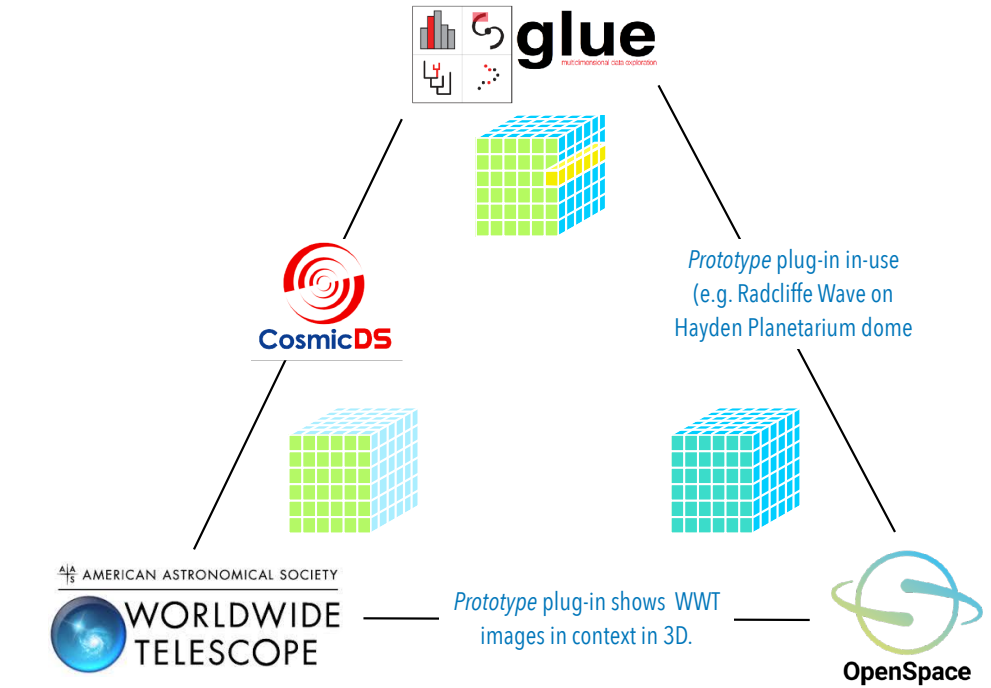




Bialy et al. 2021

GOAL = DDD in *all* of the Local Milky Way *as we know it*

***great* 1D, 2D** and **3D** data manipulation, flexible architecture facilitating plug-ins, data transfer, and interactive data exploration; "glupyter" flavor runs in web pages

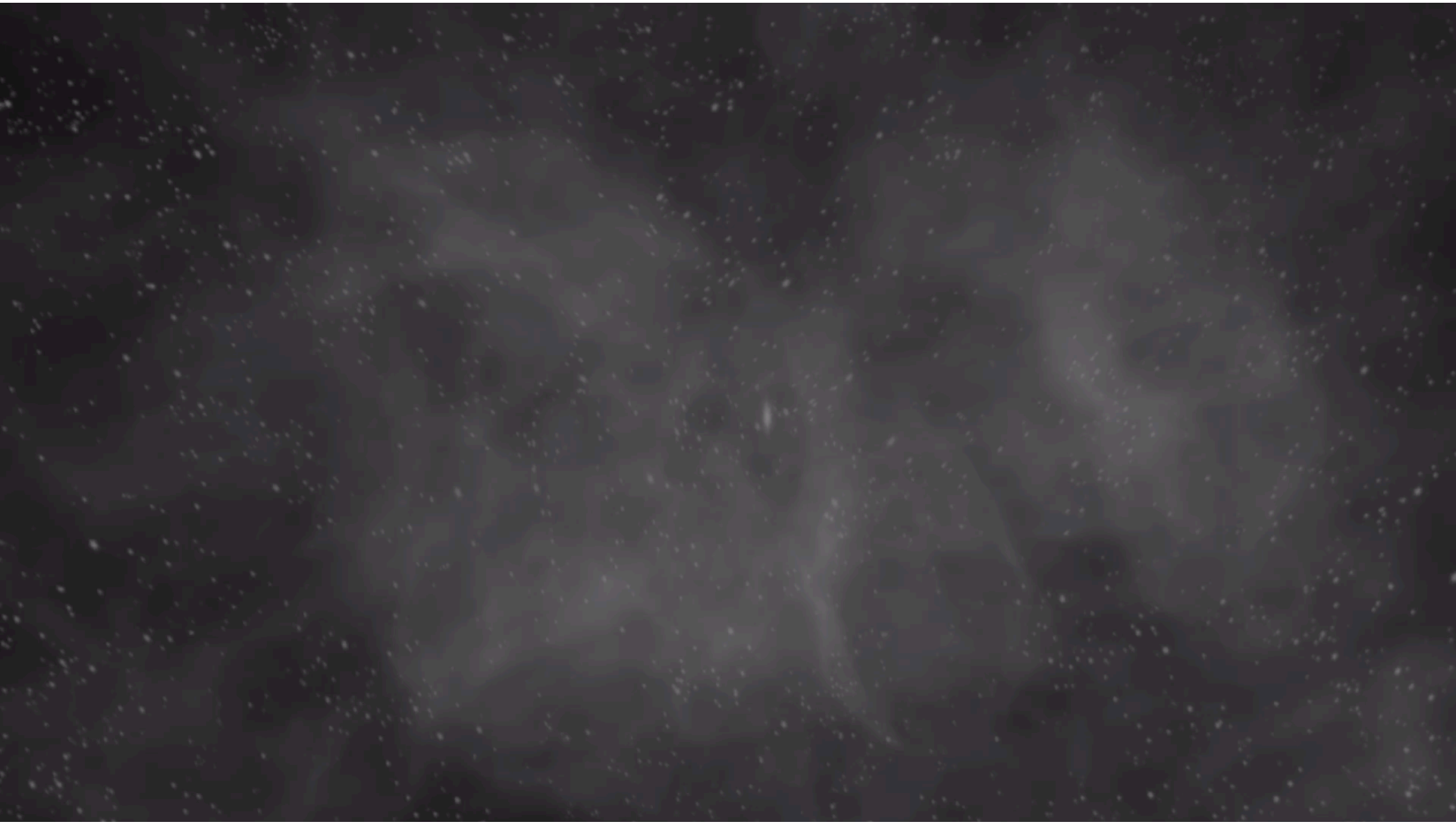


***great* 2D** object and all-sky images
limited 3D functionality

limited 2D all-sky images
***great* 3D** functionality

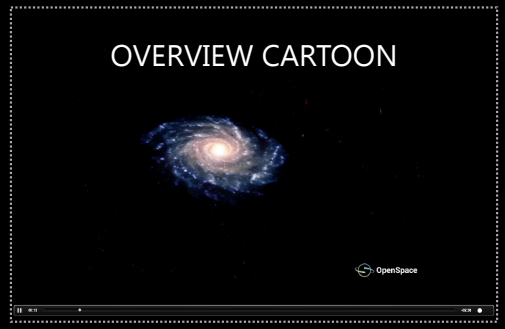
The "Perseus-Taurus Superbubble"
a demo of the need for 2D-3D contextualization functionality

This video was composited using the WWT and OpenSpace, making some use of prototype plug-ins, but 2D and 3D imagery was aligned manually by experts. As a generalizable STEM concept, it explains the deceptive "forced perspective" made possible in when objects at very different distances, in 3D, appear to touch in 2D.

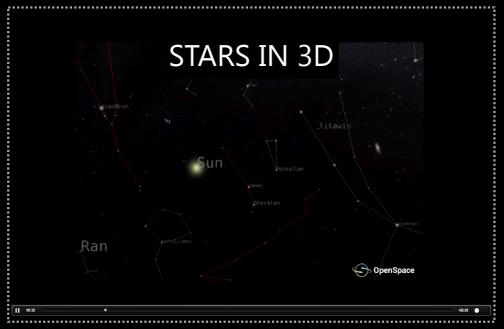


VISUALIZATION FEATURES

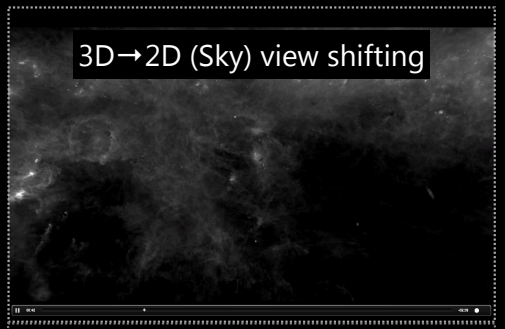
OVERVIEW CARTOON



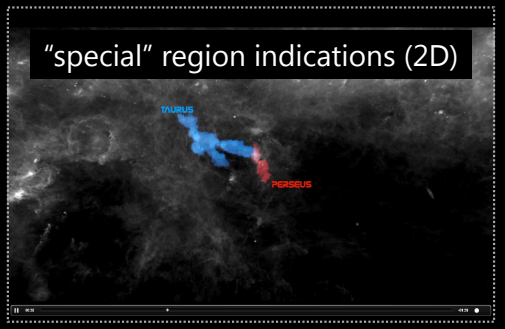
STARS IN 3D



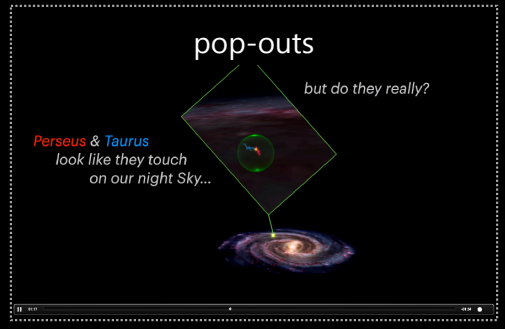
3D→2D (Sky) view shifting



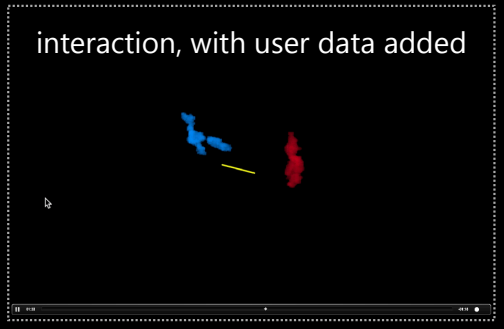
"special" region indications (2D)



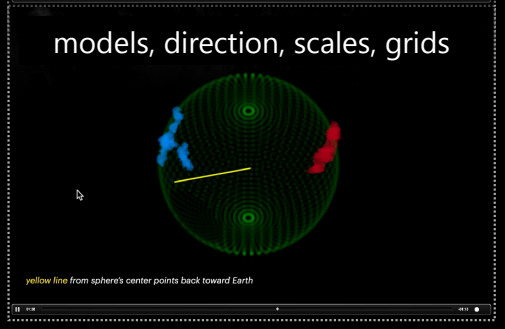
pop-outs



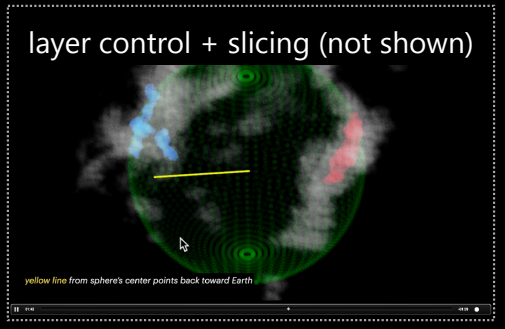
interaction, with user data added



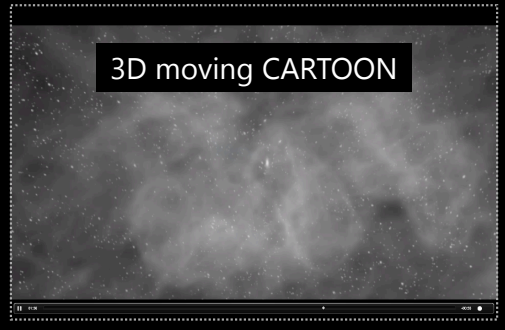
models, direction, scales, grids



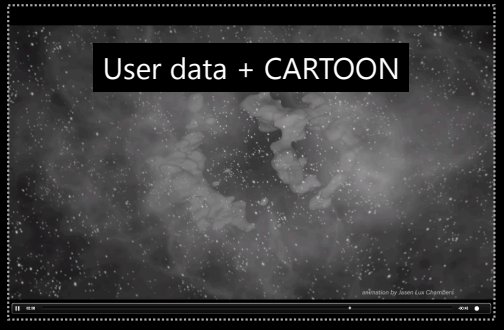
layer control + slicing (not shown)



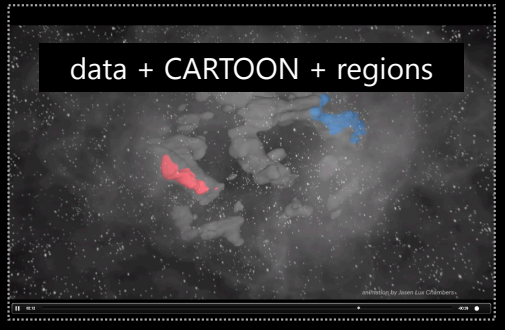
3D moving CARTOON



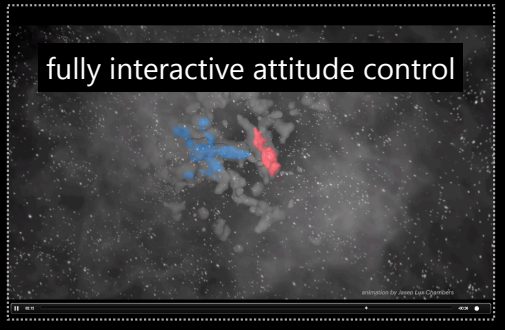
User data + CARTOON



data + CARTOON + regions



fully interactive attitude control



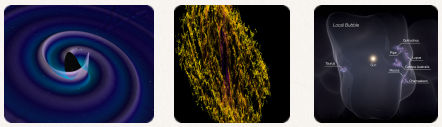
Opinion

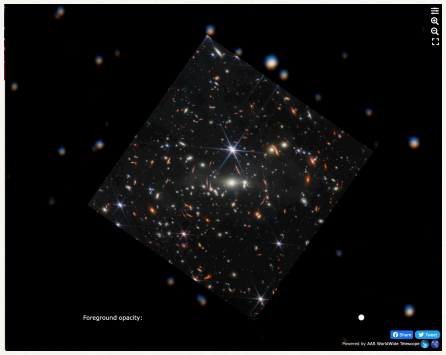
The New Universe

MEMPHIS, SUNDAY OCTOBER 23, 2022

WHAT DO EXPENSIVE NEW TELESCOPES DO FOR HUMANITY TODAY?

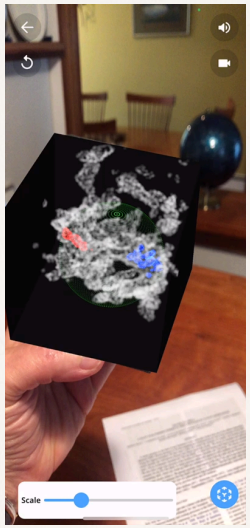
Are mega-projects like ALMA, LIGO, JWST, and Gaia worth the billions?





ARE COMPUTERS THE NEW TELESCOPES?

New galaxies in-silico, the early Universe without physics, and new stars forming in your hand.

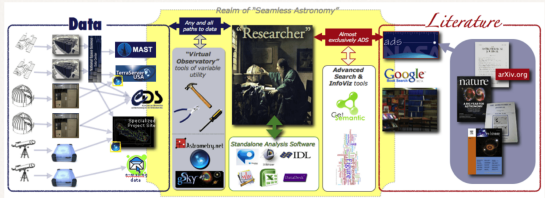


IS ASTROPHYSICS BEING (RE)ORGANIZED?

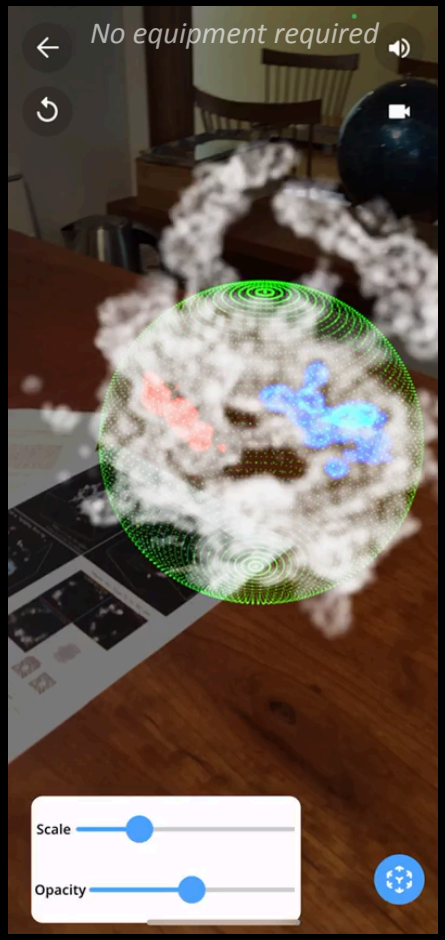
Lone stargazers are a rarer and rarer breed in professional astronomy. Teams and data scientists seem the way of the future, and tools that talk to each other are essential.

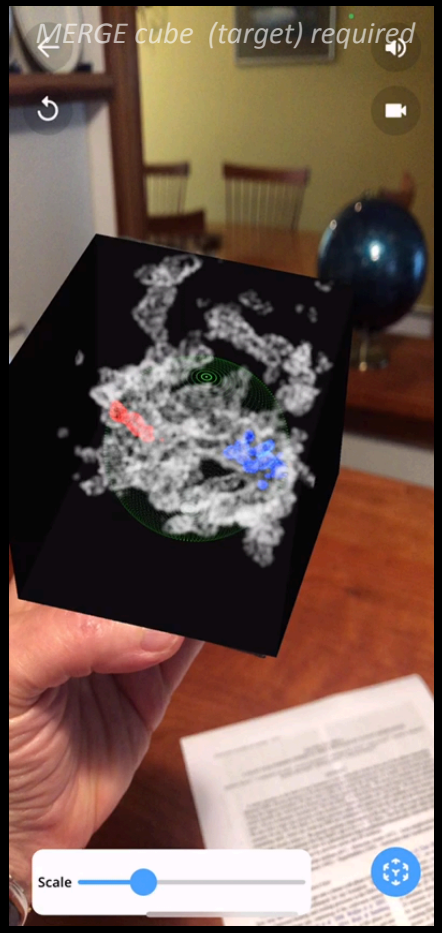
SEAMLESS ASTRONOMY

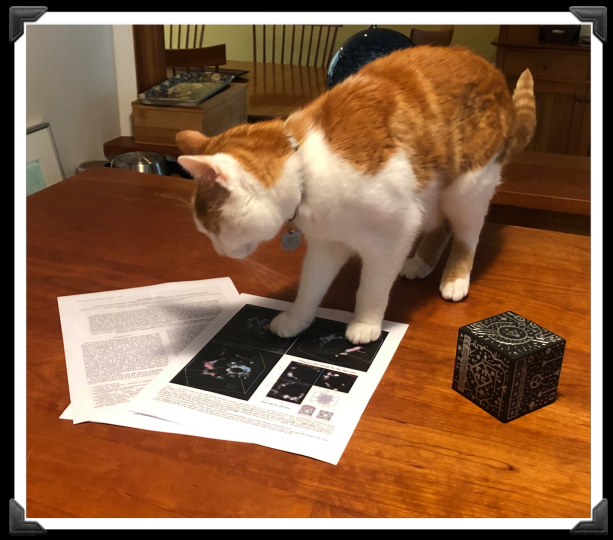
Linking scientific data, publications, and communities



Editor: Alyssa Goodman, Center for Astrophysics | Harvard & Smithsonian, @AlyssaAGoodman







Augmented reality figures from Bialy et al. 2021.

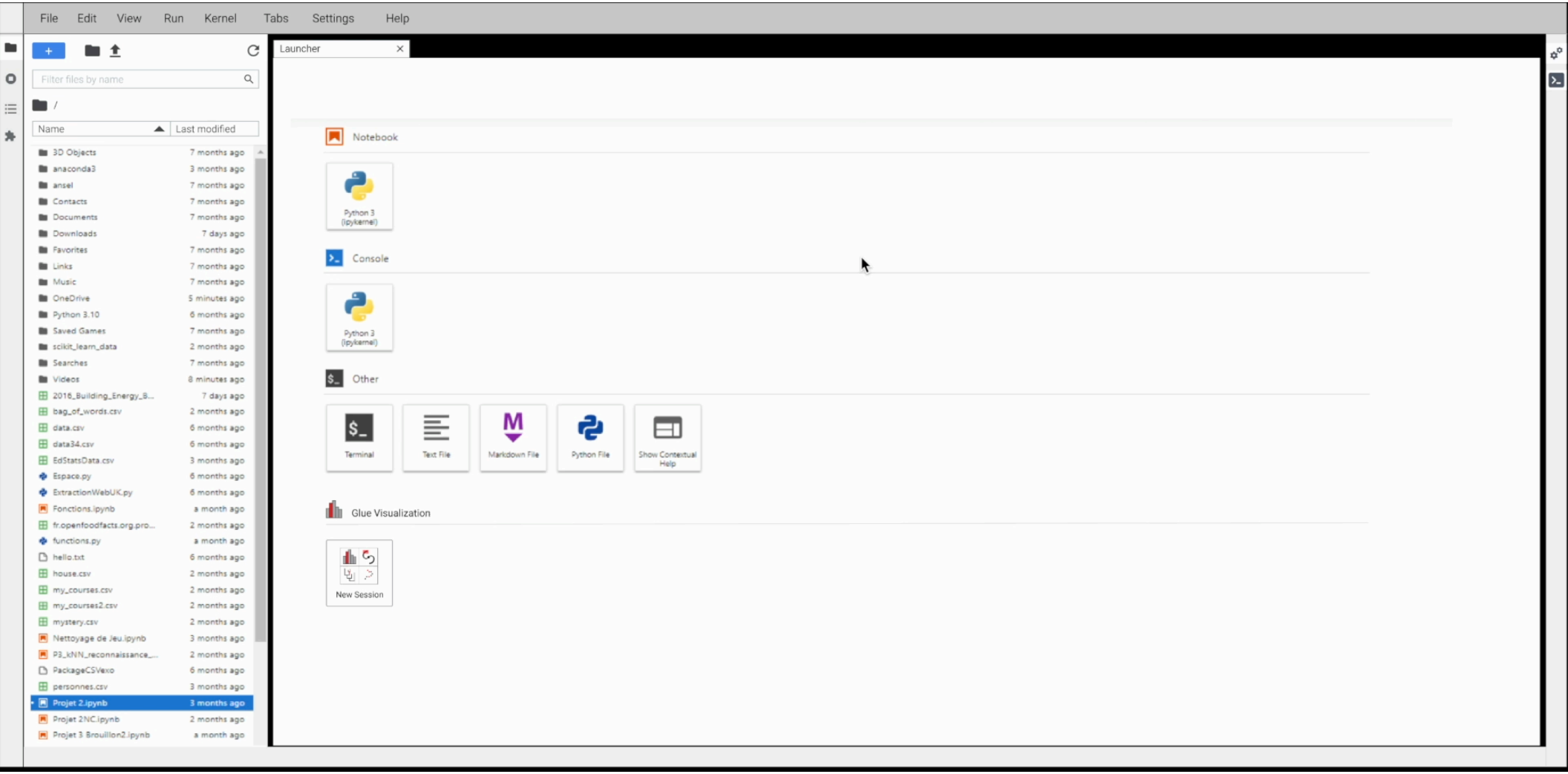
Now funded to make "AR for all" by NSF

Exploring High-Dimensional Data in Astronomy, Genomics, and beyond, using glue

Coming in 2023, **glue** *in* jupyter lab *v.0.1*

glupyter
multidimensional data exploration

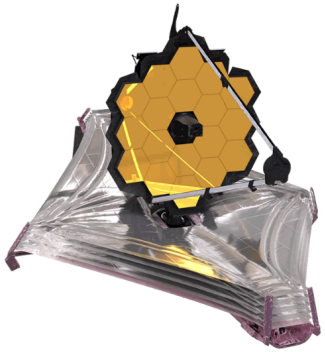
thanks to NSF & GORDON AND BETTY MOORE FOUNDATION



Cousins of glupyter

multidimensional data exploration

gluesolutions.io/the-software/glupyter



Quick insights for Images, Spectra

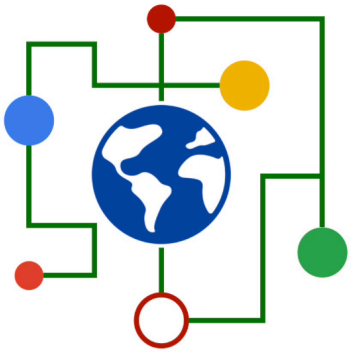
JDAViz

includes: ImViz, CubeViz, SpecViz, MOSViz

Sponsor: NASA, James Webb Space Telescope

Read more (blog post at 10QViz.org)...

GitHub



Open-Source GIS Data Exploration

SAVE

Search-Analysis-Visualization-Environment

Sponsors: Harvard+Google Data+Climate

Read more at Data+Climate site...

GitHub

CosmicDS

Data Science Education

Cosmic Data Stories

Sponsor: NASA, Science Activation Program (funded proposal)

Read more at CosmicDS website...

GitHub

Exploring High-Dimensional Data in Astronomy, Genomics, and beyond, using glue


glue
genes

Exploring High-Dimensional Data in Astronomy, Genomics, and beyond, using glue

Alyssa Goodman
Robert Wheeler Willson Professor, Astronomy, Harvard University
& President, glue solutions, inc.

Jonathan Foster
Chief Technology Officer, glue solutions, inc.

AstroMed

AMERICAN ASTRONOMICAL SOCIETY WORLDWIDE TELESCOPE

glue
multidimensional data exploration

glupyter
multidimensional data exploration

glue genes

NASA NSF HARVARD GORDON AND BETTY MOORE FOUNDATION The Jackson Laboratory glue solutions inc.

glueviz.org gluesolutions.io

tinyurl.com/dimensionsofdiscovery

agoodman@cfa.harvard.edu jfoster@gluesolutions.io

AstroMed

AMERICAN ASTRONOMICAL SOCIETY WORLDWIDE TELESCOPE

glue multidimensional data exploration

glupyter multidimensional data exploration

glue genes