glueing together the Universe

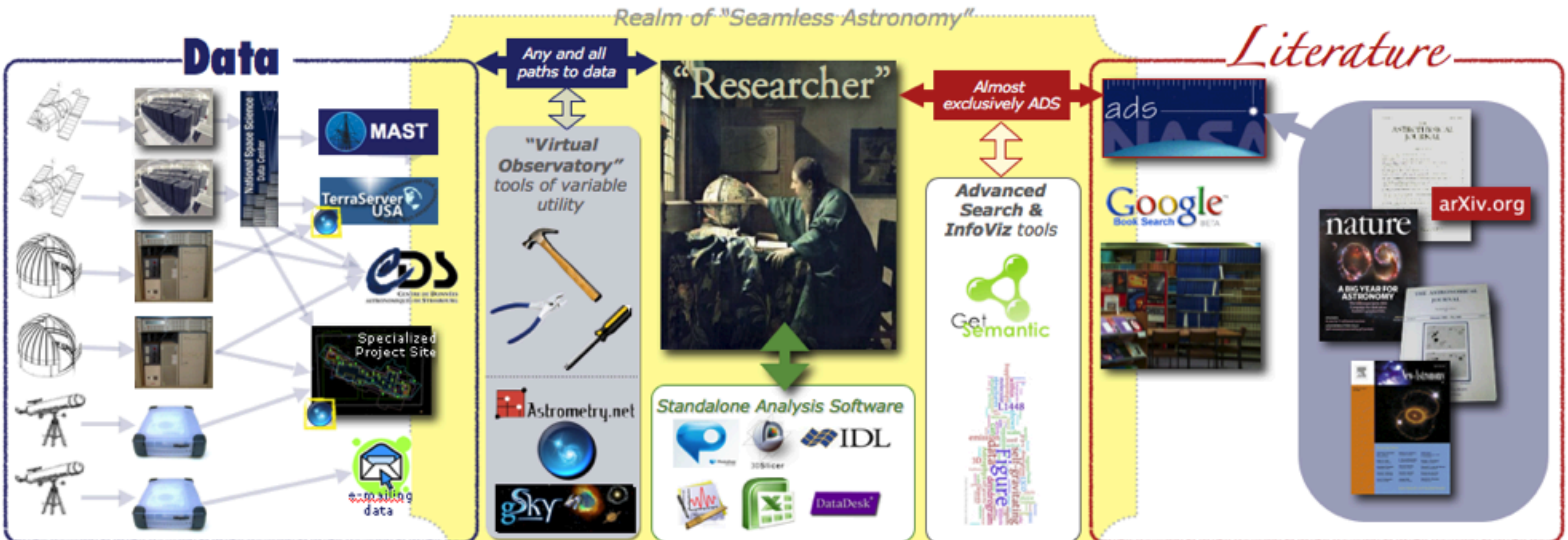
multidimensional data exploration

and the Earth

Alyssa A. Goodman • Harvard-Smithsonian Center for Astrophysics • @aagie

SEAMLESS ASTRONOMY

Linking scientific data, publications, and communities

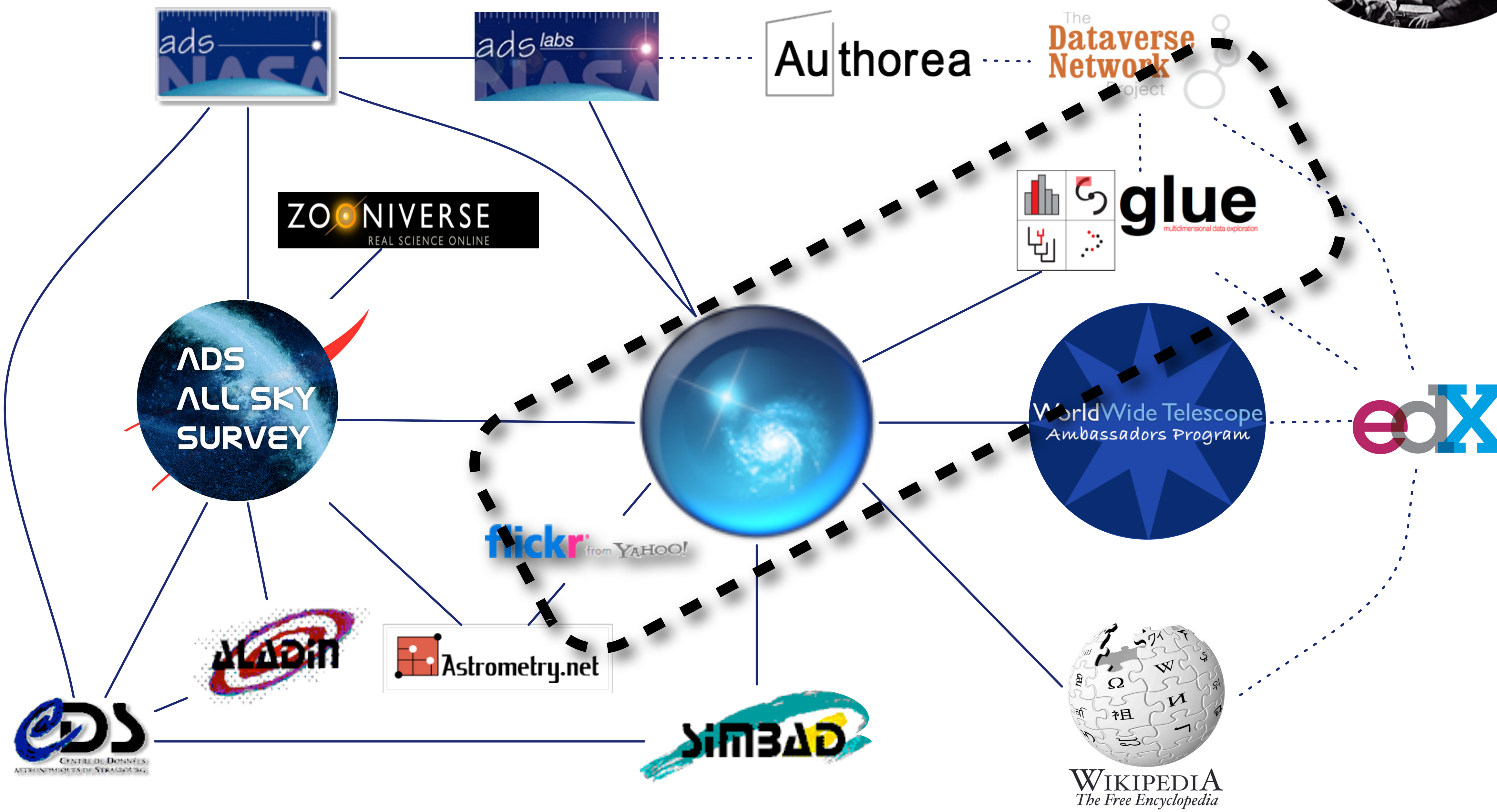


projects.iq.harvard.edu/seamlessastronomy

Alberto Accomazzi, Christopher Beaumont, Douglas Burke, Raffaele D'Abrusco, Rahul Davé, Christopher Erdmann, Pepi Fabbiano, Alyssa Goodman, Edwin Henneken, Jay Luker, Gus Muench, Michael Kurtz, Max Lu, Victoria Mittelbach, Alberto Pepe, Arnold Rots, Patricia Udomprasert (Harvard-Smithsonian CfA); Mercé Crosas (Harvard Institute for Quantitative Social Science); Christine Borgman (UCLA); Jonathan Fay & Curtis Wong (Microsoft Research); Alberto Conti (Space Telescope Science Institute)

SEAMLESS ASTRONOMY

Linking scientific data, publications, and communities



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

Supported by Microsoft Research NSF NASA Smithsonian

Evolution since the Revolution

ASTROPHYSICAL JOURNAL

AN INTERNATIONAL REVIEW OF SPECTROSCOPY AND ASTRONOMICAL PHYSICS

VOLUME I — JANUARY 1895 — NUMBER 1

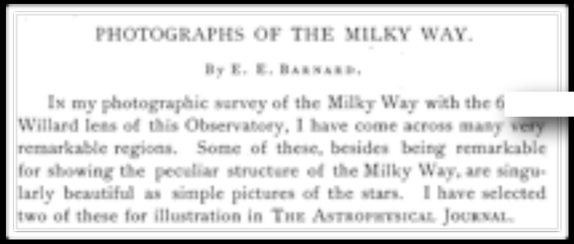
ON THE CONDITIONS WHICH AFFECT THE SPECTRO-PHOTOGRAPHY OF THE SUN.

By Albert A. Michelson.

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

Evidently the process is not limited to the photography of the prominences, but extends to all other peculiarities of structure which emit radiations of approximately constant wave-length; and the efficiency of the method depends very largely upon the *contrast* which can be obtained by the greater enfeeble-

1665 ..230 yr... 1895 ...114 yr... 2009 ...4 yr... 2013

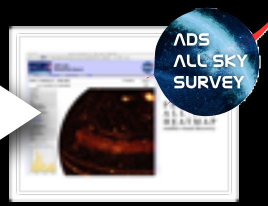


PHOTOGRAPHS OF THE MILKY WAY.

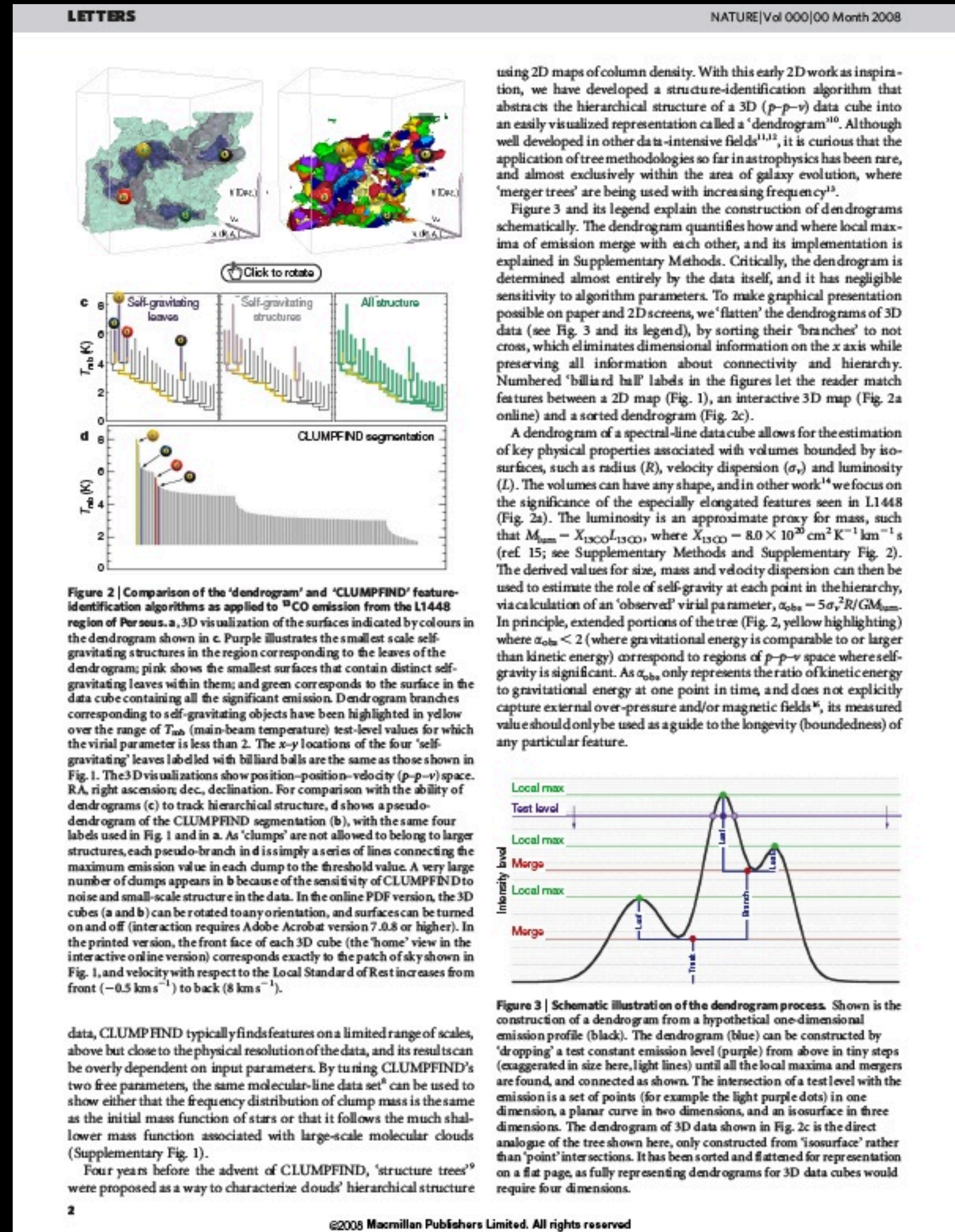
By E. E. Barnard.

In my photographic survey of the Milky Way with the 6 Willard lens of this Observatory, I have come across many very remarkable regions. Some of these, besides being remarkable for showing the peculiar structure of the Milky Way, are singularly beautiful as simple pictures of the stars. I have selected two of these for illustration in The Astrophysical Journal.

ads NASA → flickr from YAHOO! Astrometry.net → ADS ALL SKY SURVEY



3D PDF



Click to rotate

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).

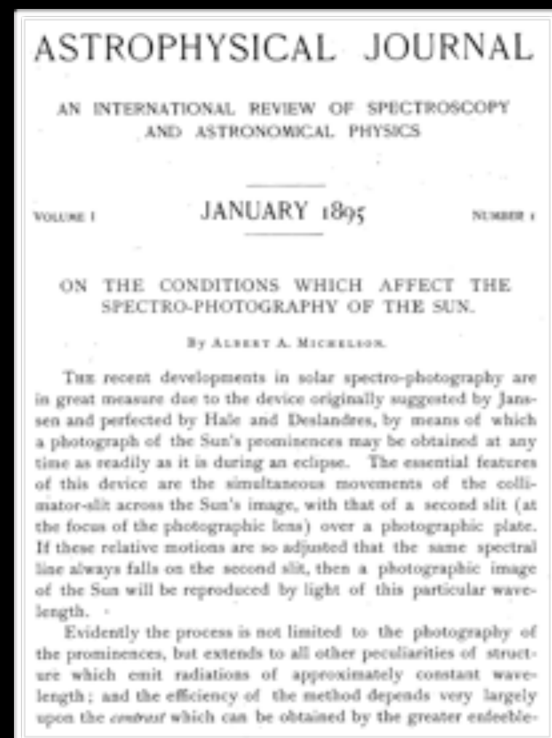
Four years before the advent of CLUMPFIND, 'structure trees'[9] were proposed as a way to characterize clouds' hierarchical structure using 2D maps of column density. With this early 2D work as inspiration, we have developed a structure-identification algorithm that abstracts the hierarchical structure of a 3D (p–p–v) data cube into an easily visualized representation called a 'dendrogram'[10]. Although well developed in other data-intensive fields[11,12], it is curious that the application of tree methodologies so far in astrophysics has been rare, and almost exclusively within the area of galaxy evolution, where 'merger trees' are being used with increasing frequency[13].

Figure 3 and its legend explain the construction of dendrograms schematically. The dendrogram quantifies how and where local maxima of emission merge with each other, and its implementation is explained in Supplementary Methods. Critically, the dendrogram is determined almost entirely by the data itself, and it has negligible sensitivity to algorithm parameters. To make graphical presentation possible on paper and 2D screens, we 'flatten' the dendrograms of 3D data (see Fig. 3 and its legend), by sorting their 'branches' to not cross, which eliminates dimensional information on the x axis while preserving all information about connectivity and hierarchy. Numbered 'billiard ball' labels in the figures let the reader match features between a 2D map (Fig. 1), an interactive 3D map (Fig. 2a online) and a sorted dendrogram (Fig. 2c).

A dendrogram of a spectral-line data cube allows for the estimation of key physical properties associated with volumes bounded by isosurfaces, such as radius (R), velocity dispersion (σ_v) and luminosity (L). The volumes can have any shape, and in other work[14] we focus on the significance of the especially elongated features seen in L1448 (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.

Evolution since the Revolution



ASTROPHYSICAL JOURNAL

AN INTERNATIONAL REVIEW OF SPECTROSCOPY AND ASTRONOMICAL PHYSICS

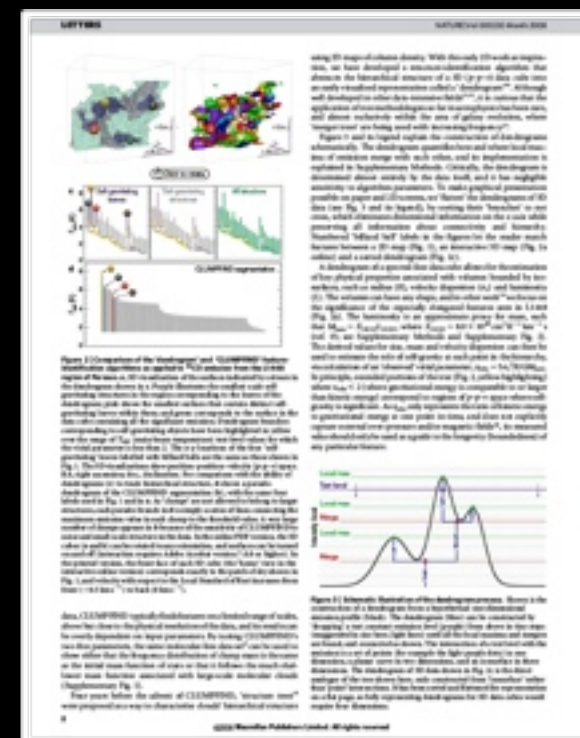
VOLUME I | JANUARY 1895 | NUMBER 1

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

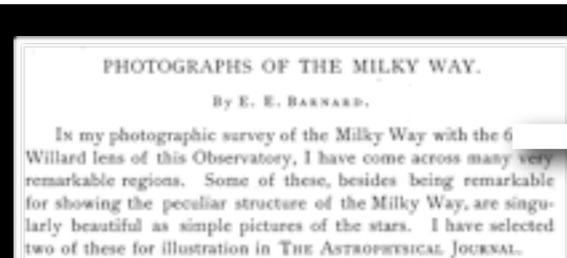
By Albert A. Michelson.

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

Evidently the process is not limited to the photography of the prominences, but extends to all other peculiarities of structure which emit radiations of approximately constant wave-length; and the efficiency of the method depends very largely upon the *contrast* which can be obtained by the greater enfeeble-



1665 ..230 yr... 1895 ...114 yr... 2009 ...4 yr... 2013



PHOTOGRAPHS OF THE MILKY WAY.

By E. E. Barnard.

In my photographic survey of the Milky Way with the 6 Willard lens of this Observatory, I have come across many very remarkable regions. Some of these, besides being remarkable for showing the peculiar structure of the Milky Way, are singularly beautiful as simple pictures of the stars. I have selected two of these for illustration in The Astrophysical Journal.

ads NASA → flickr from YAHOO! Astrometry.net →

[demo flickr-WWT]

The WorldWide Telescope+WorldWide Telescope Ambassadors Collaborations

[Demo]

Experience WWT at **worldwidetelescope.org**

glueing together the Universe

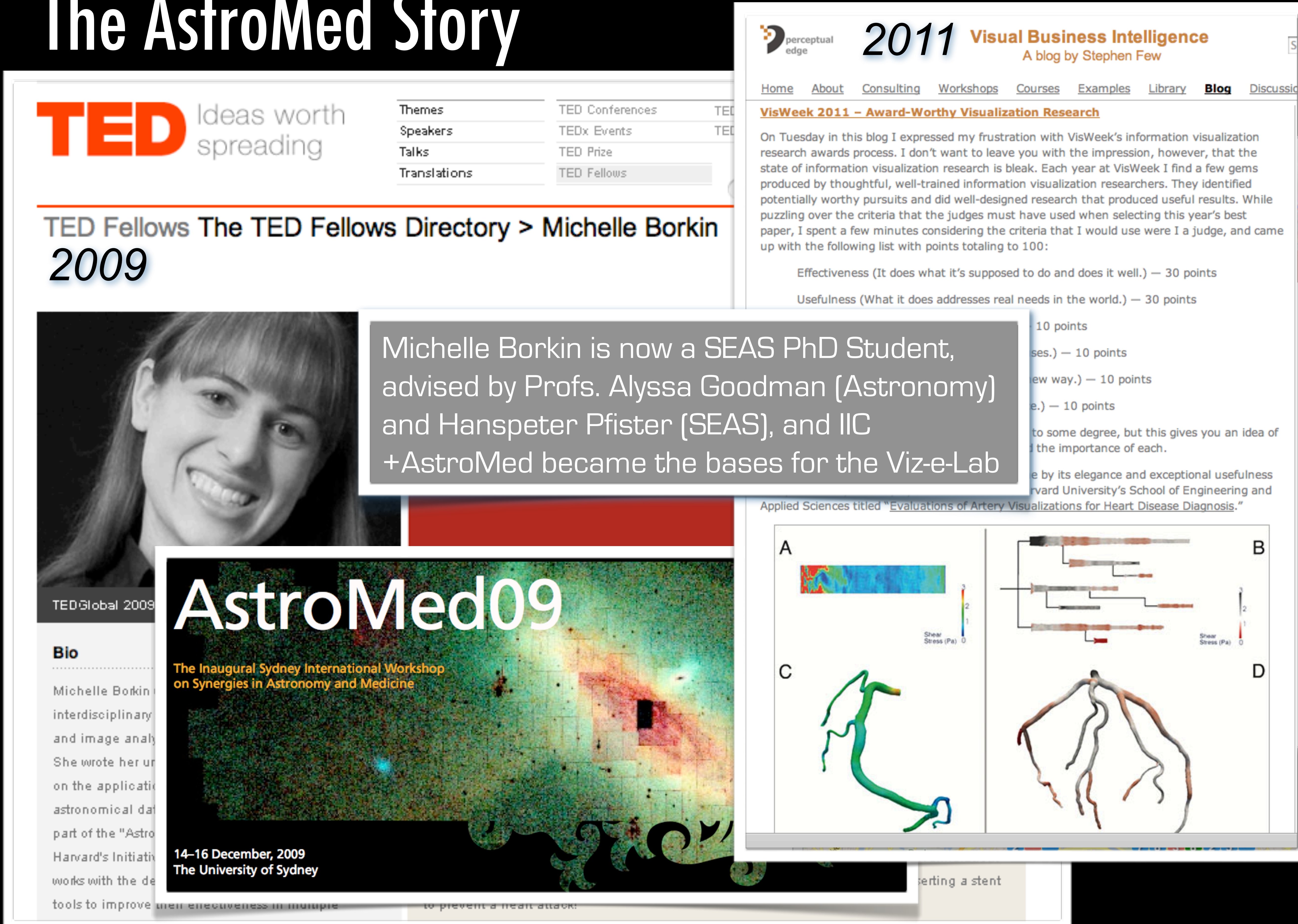
multidimensional data exploration

and the Earth

glue

multidimensional data exploration

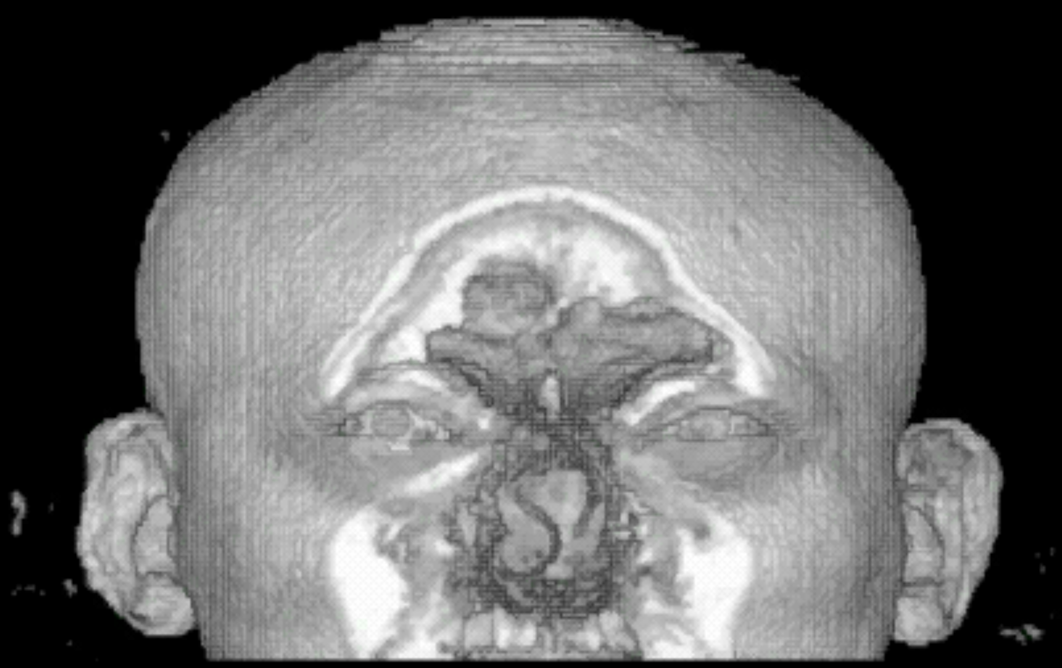
The AstroMed Story



Michelle Borkin is now a SEAS PhD Student, advised by Profs. Alyssa Goodman (Astronomy) and Hanspeter Pfister (SEAS), and IIC +AstroMed became the bases for the Viz-e-Lab

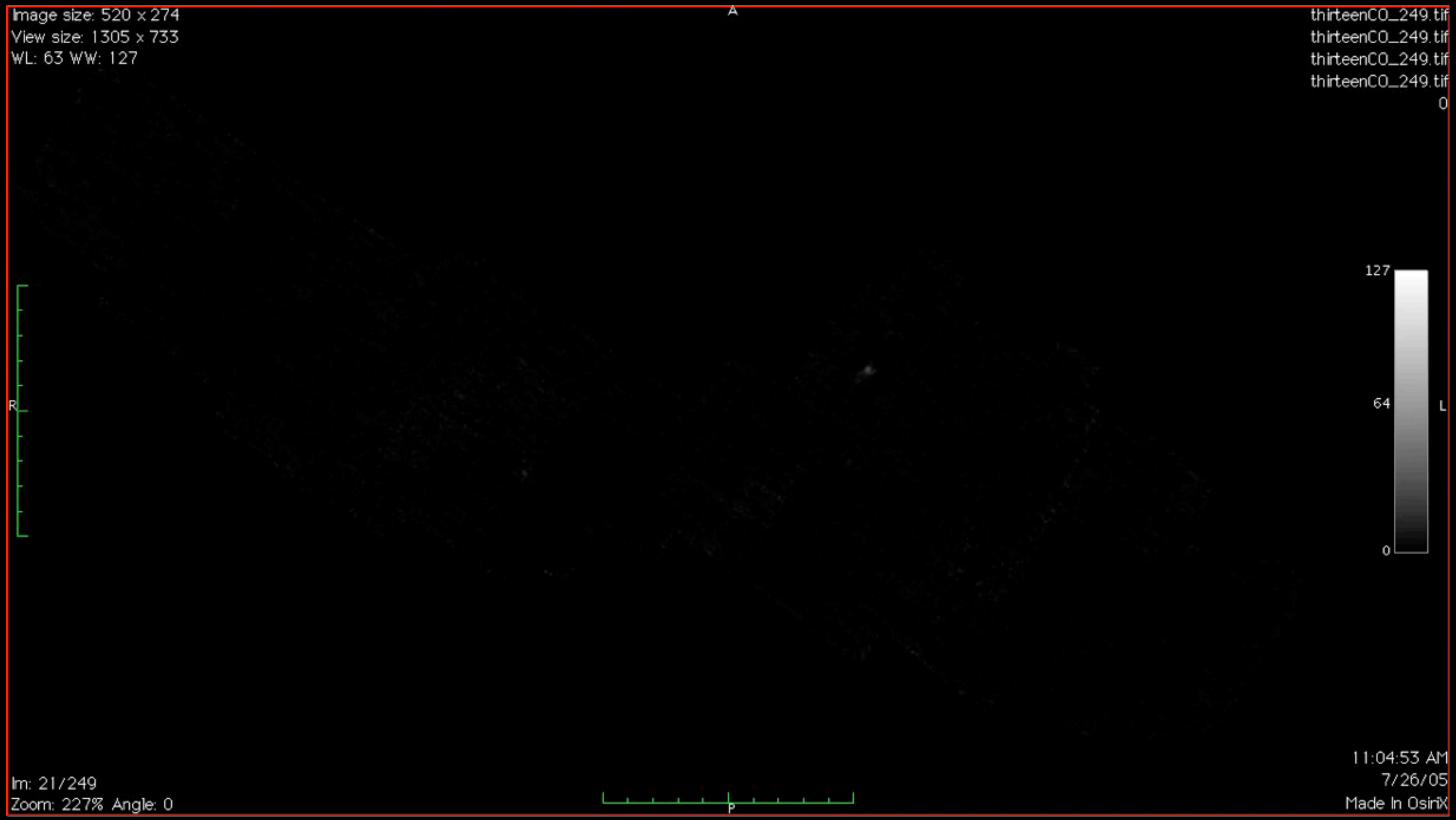
"Astronomical Medicine"

"KEITH"



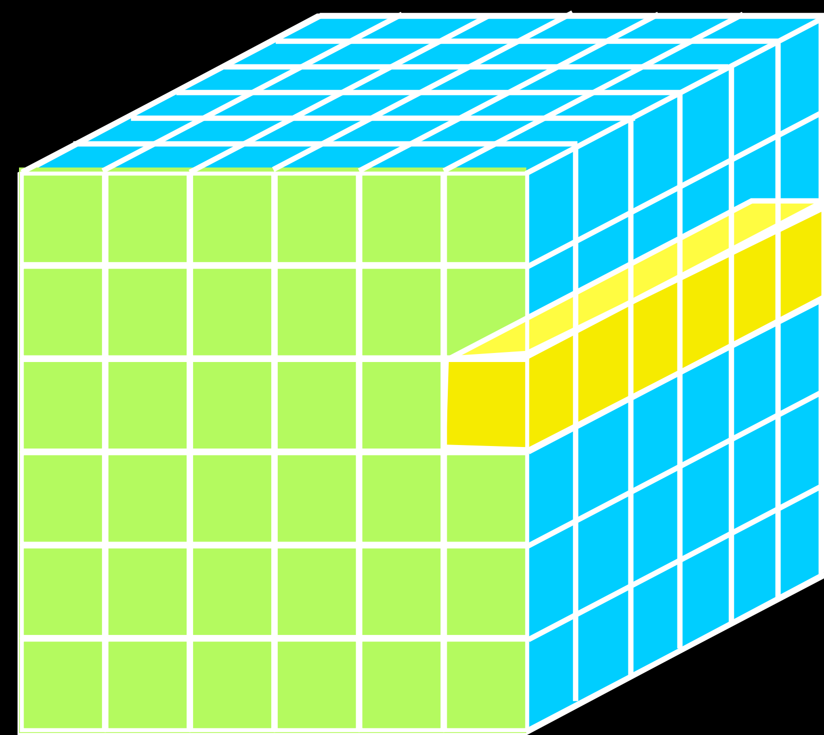
"z" is depth into head

"PERSEUS"



"z" is line-of-sight velocity

(This kind of "series of 2D slices view" is known in the Viz as "the grand tour")



GENERALLY

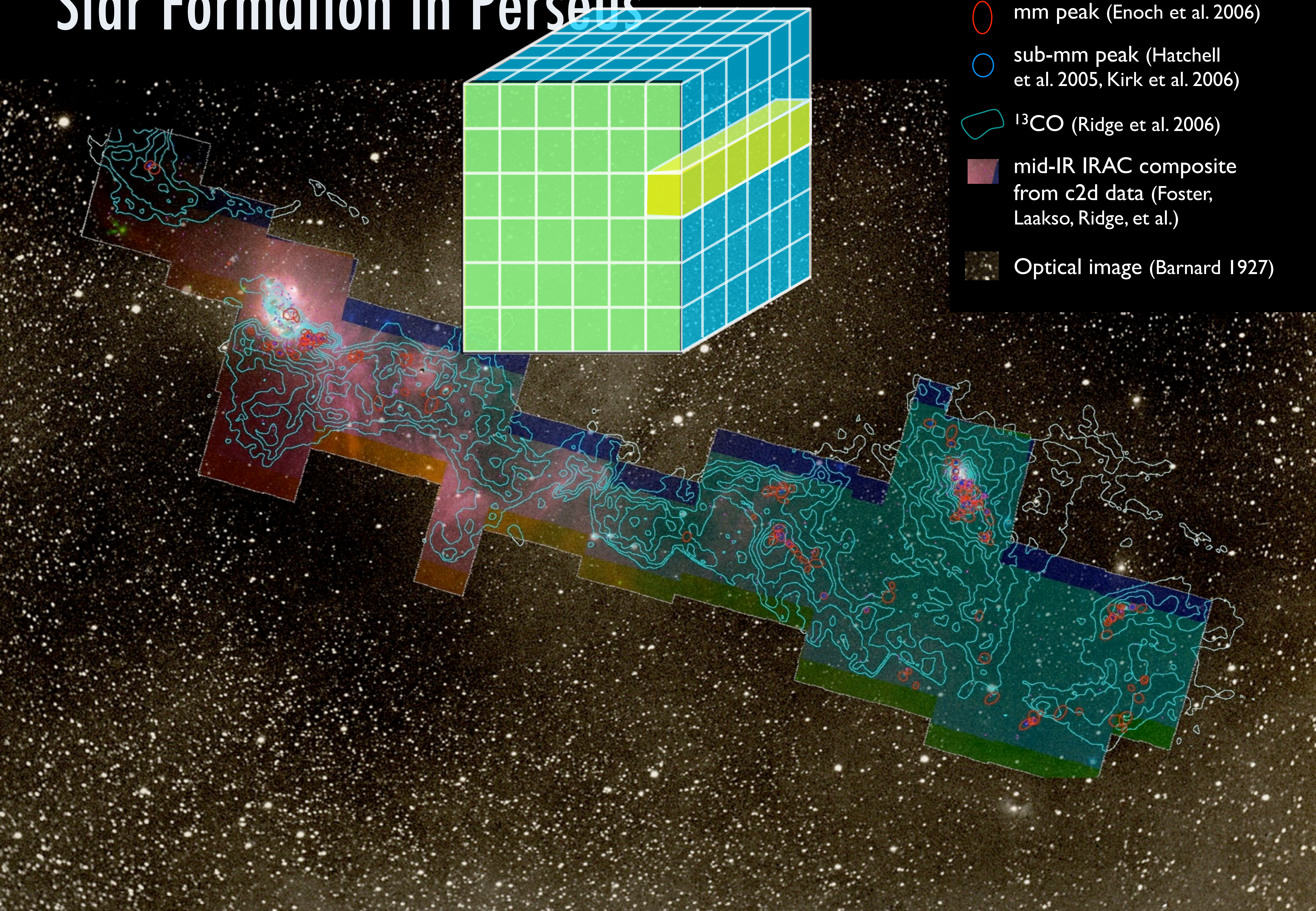
1D: Columns = “Spectra”, “SEDs” or “Time Series”

2D: Faces or Slices = “Images”

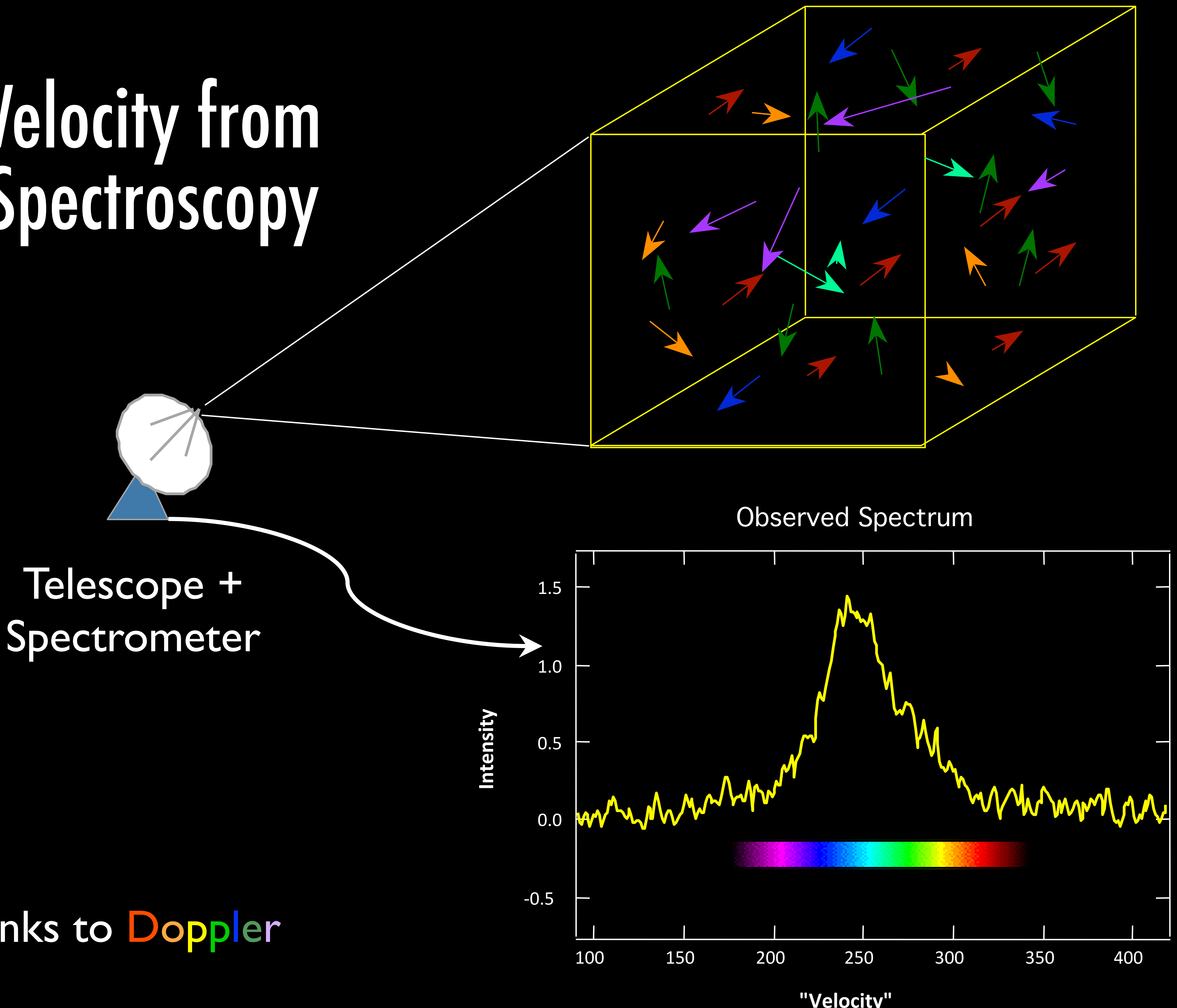
3D: Volumes = “3D Renderings”, “2D Movies”

4D: Time Series of Volumes = “3D Movies”

Star Formation in Perseus

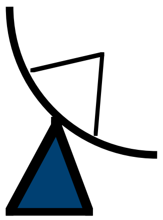


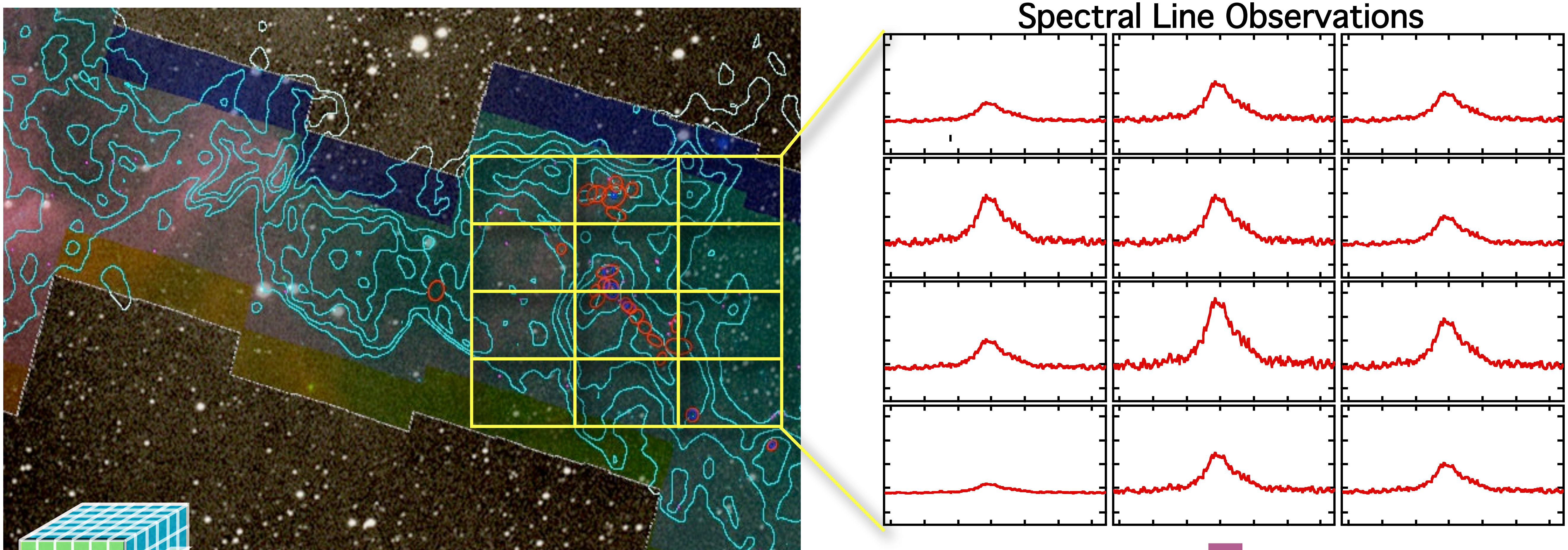
Velocity from Spectroscopy

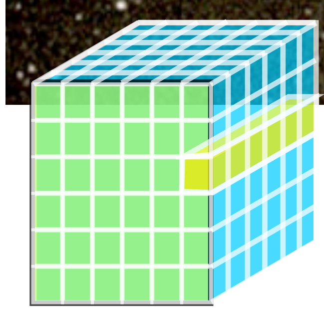


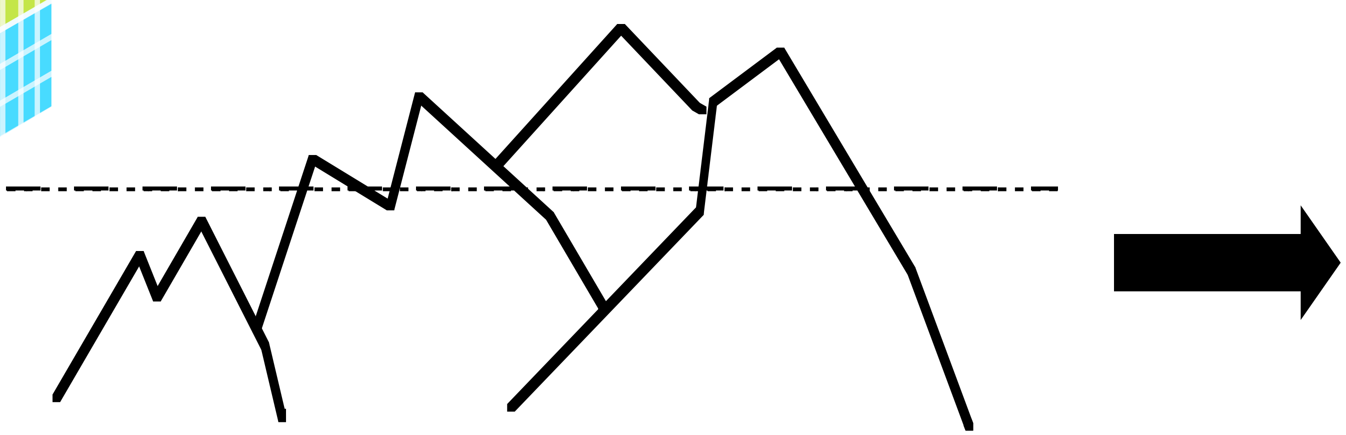
All thanks to Doppler

Spectral-Line Mapping



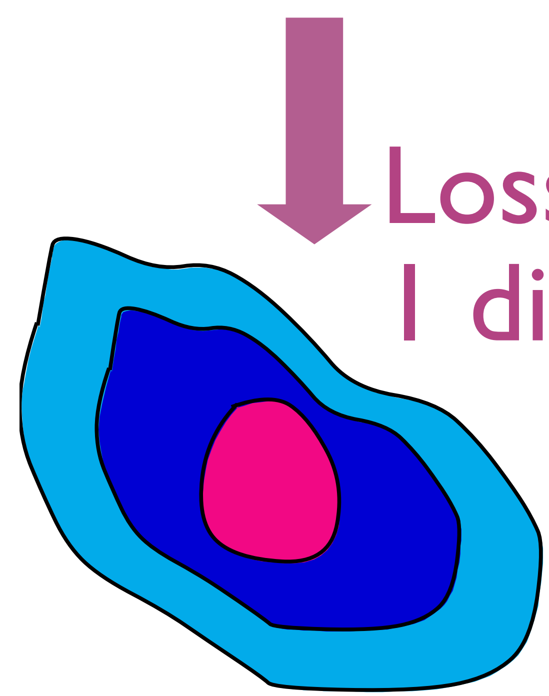






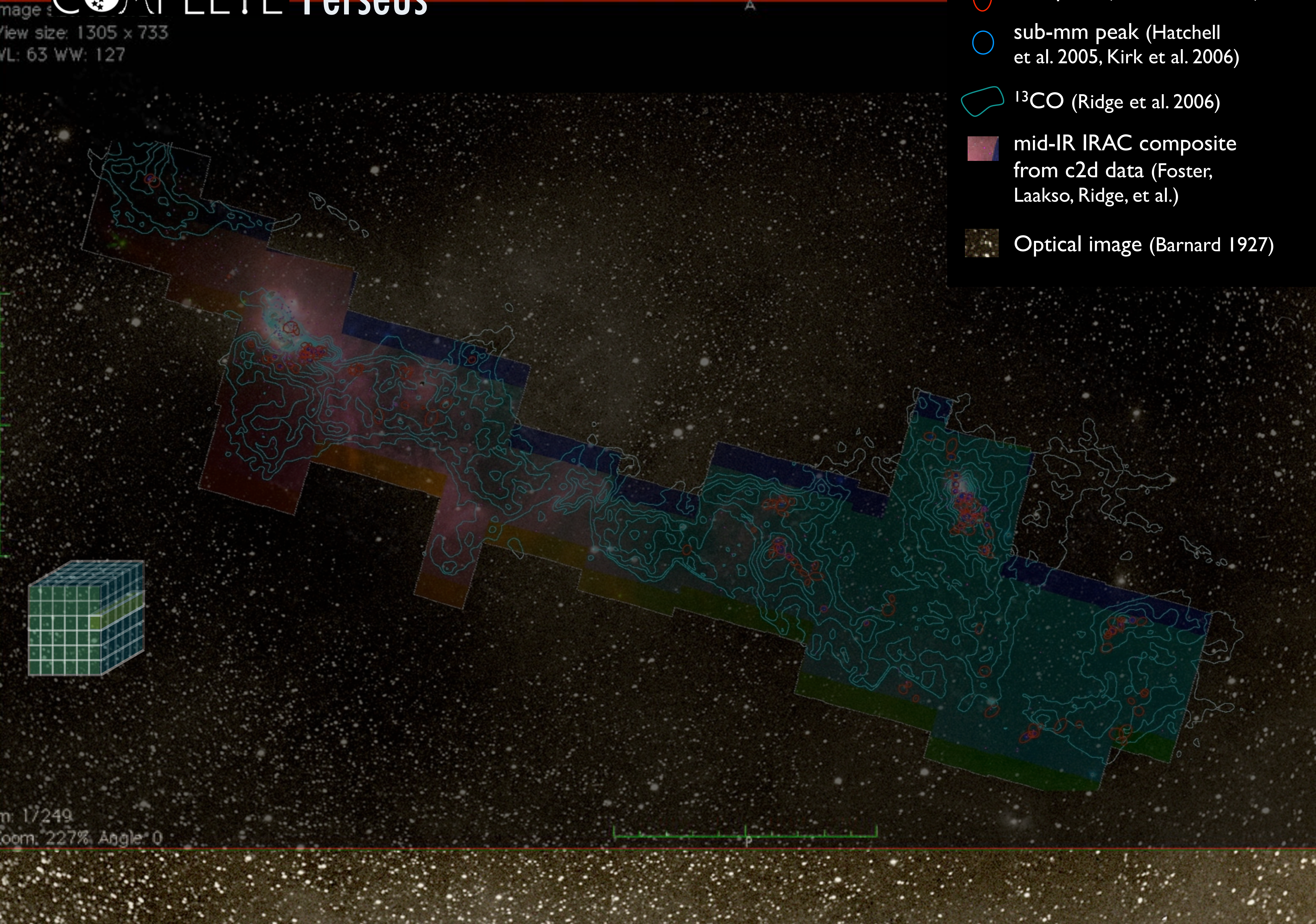
Mountain Range

No loss of information



Loss of 1 dimension

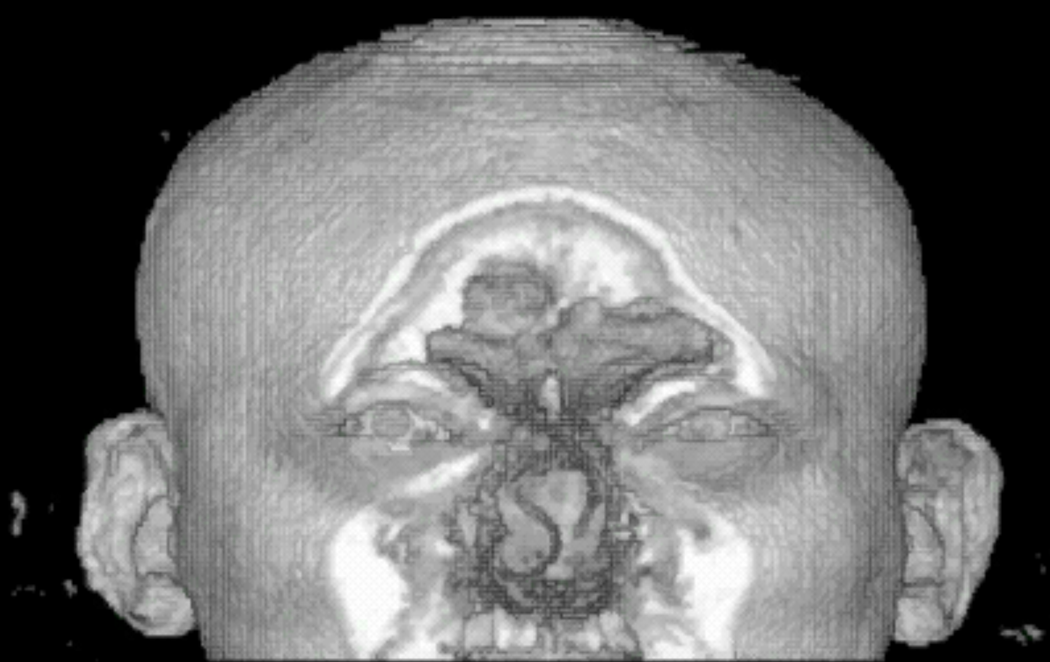
COMPLETE Perseus



- 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)

“Astronomical Medicine”

“KEITH”



“z” is depth into head

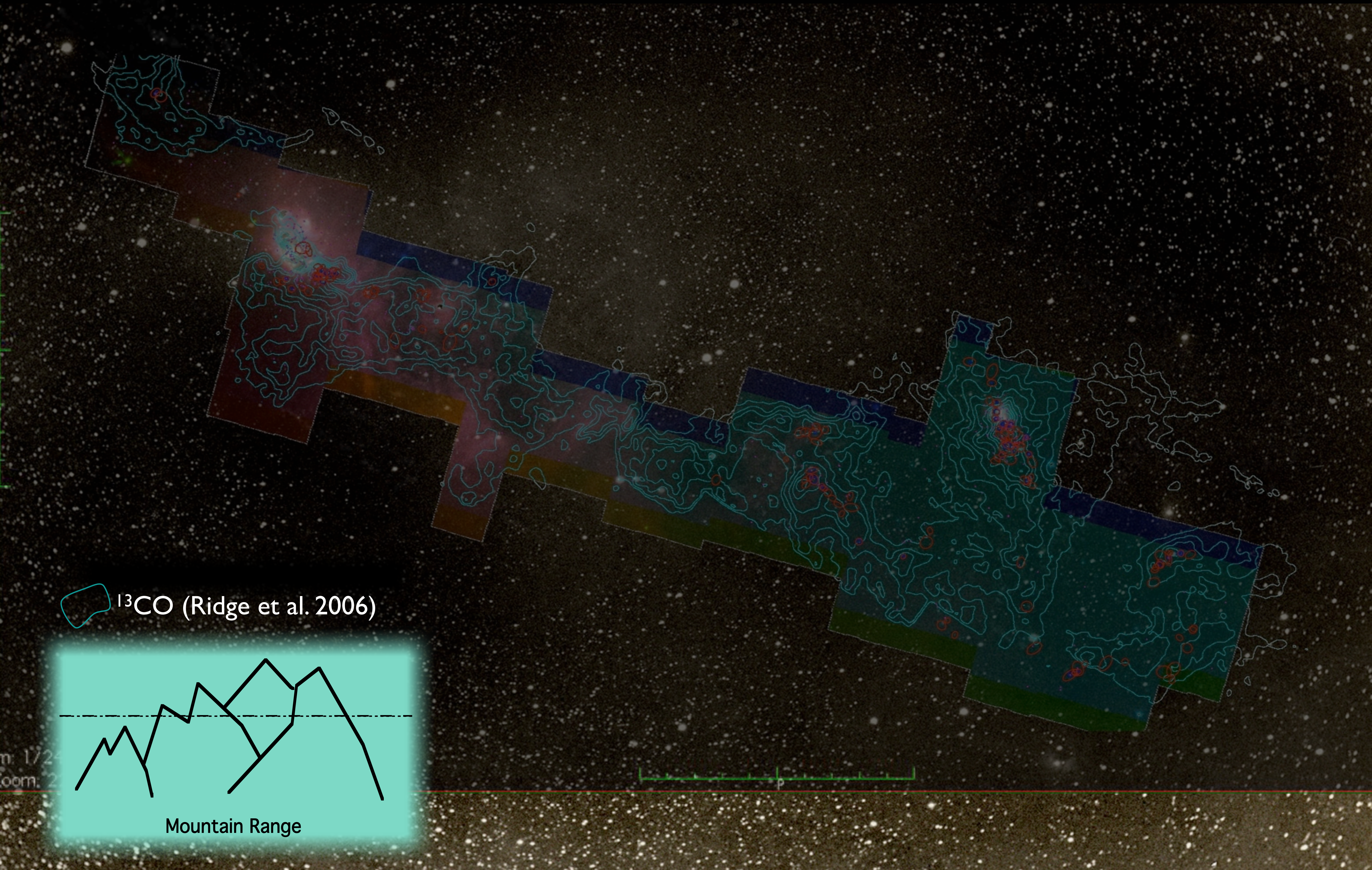
“PERSEUS”

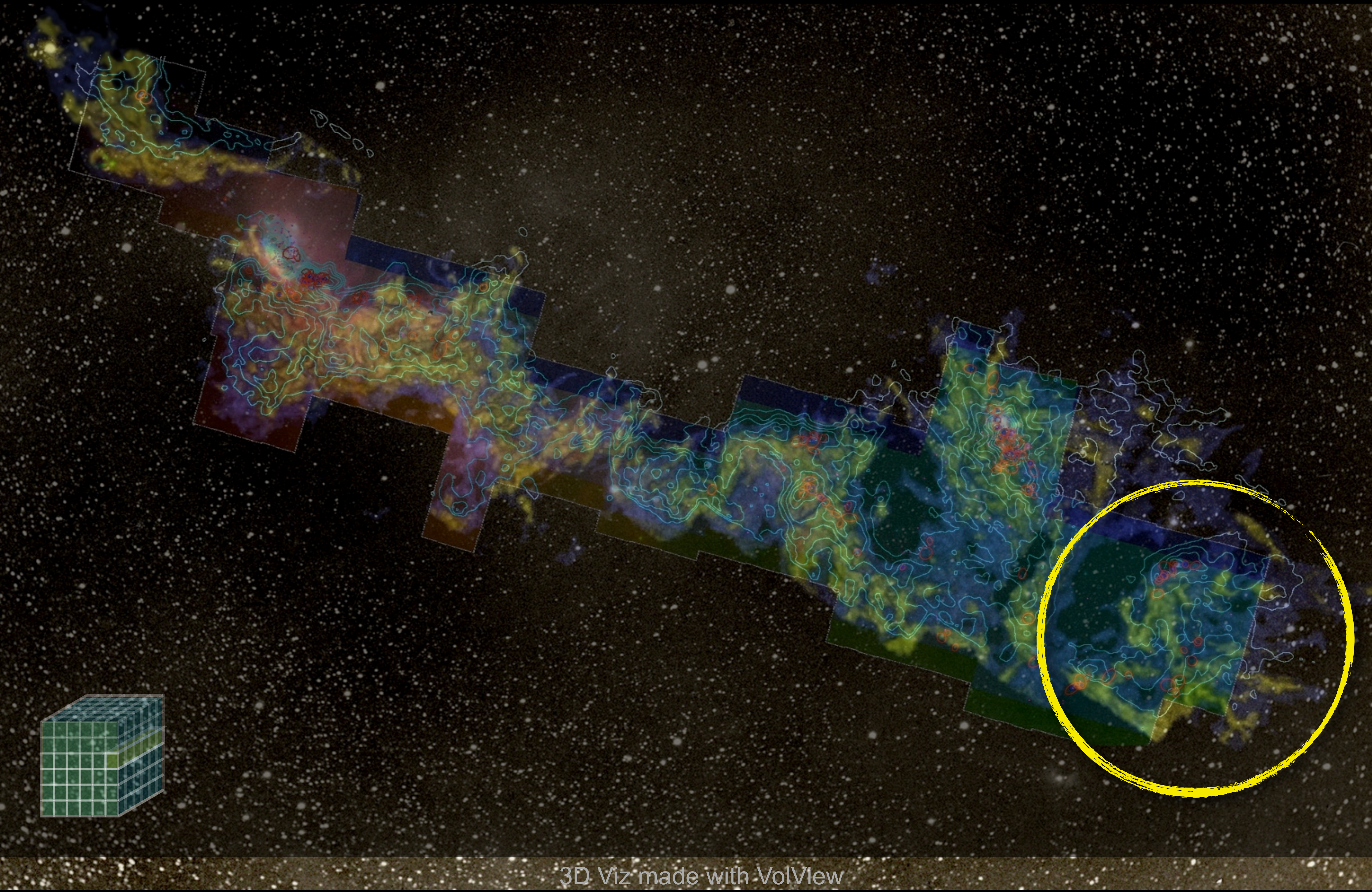
“z” is line-of-sight velocity

http://am.iic.harvard.edu/

mage size: 520 x 274
iew size: 1305 x 733
VL: 63 WW: 127

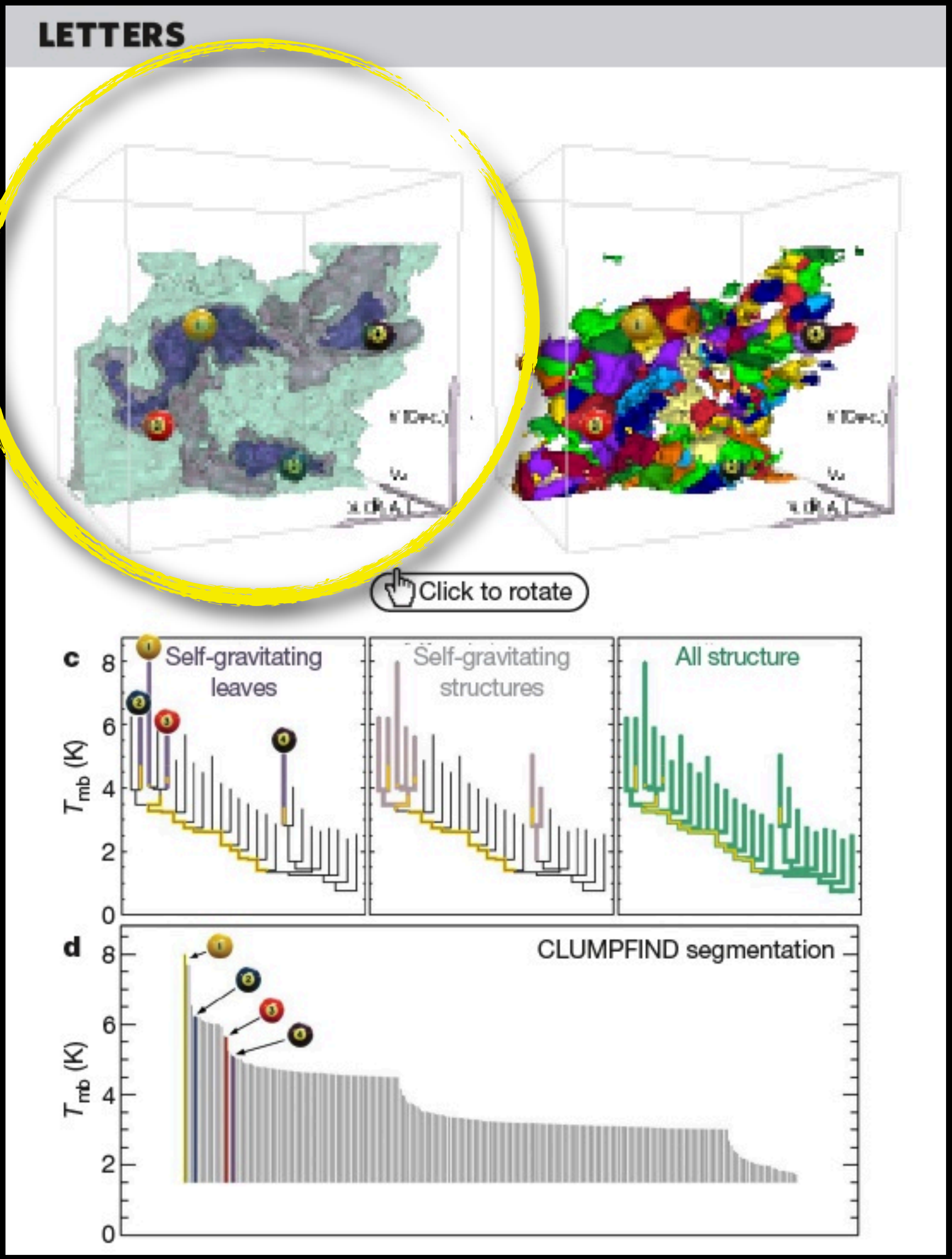
COMPLETE Perseus



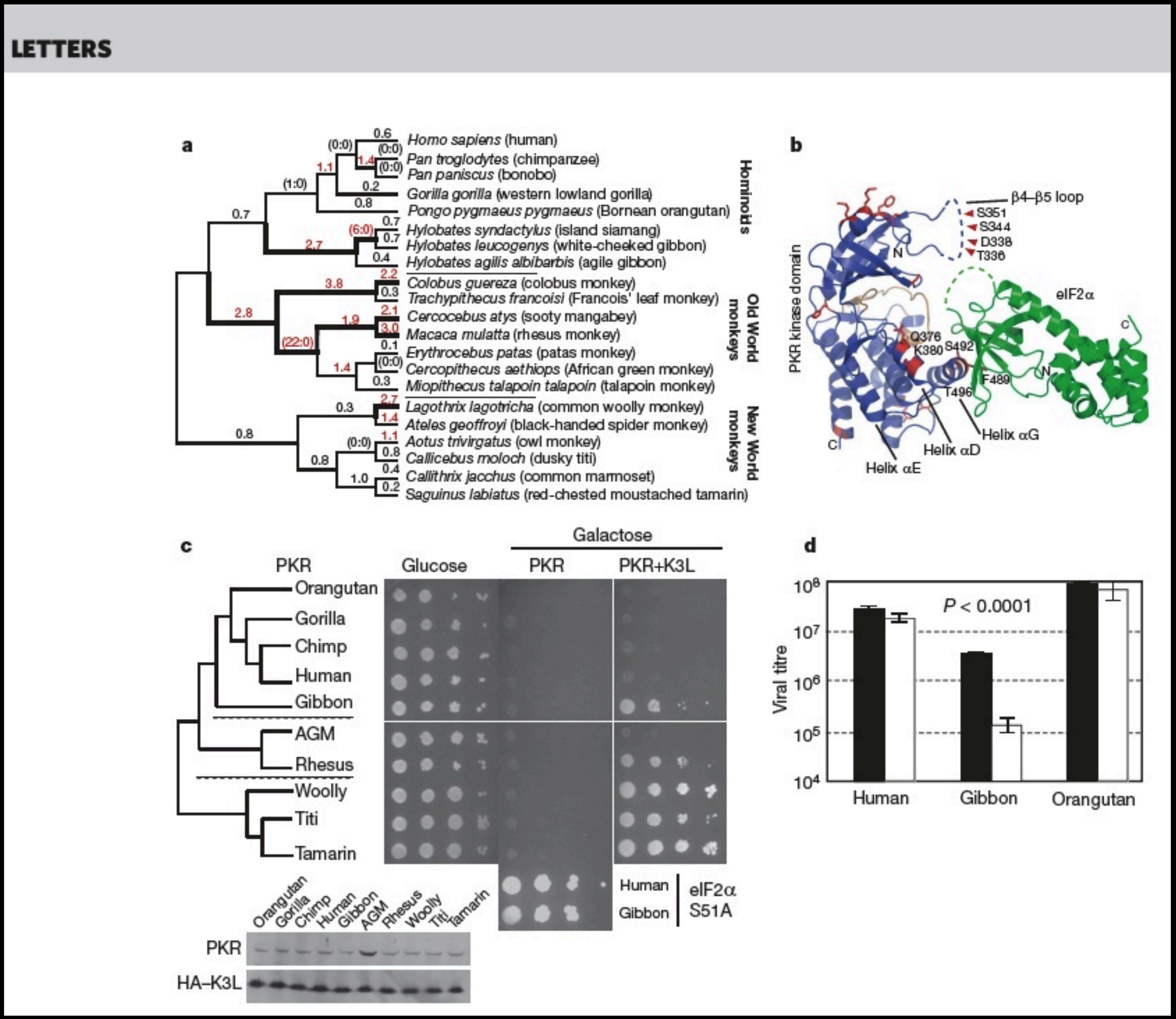

3D Viz made with VolView

“High-dimensional” or “Multivariate” Data

(Astronomy=Biology)

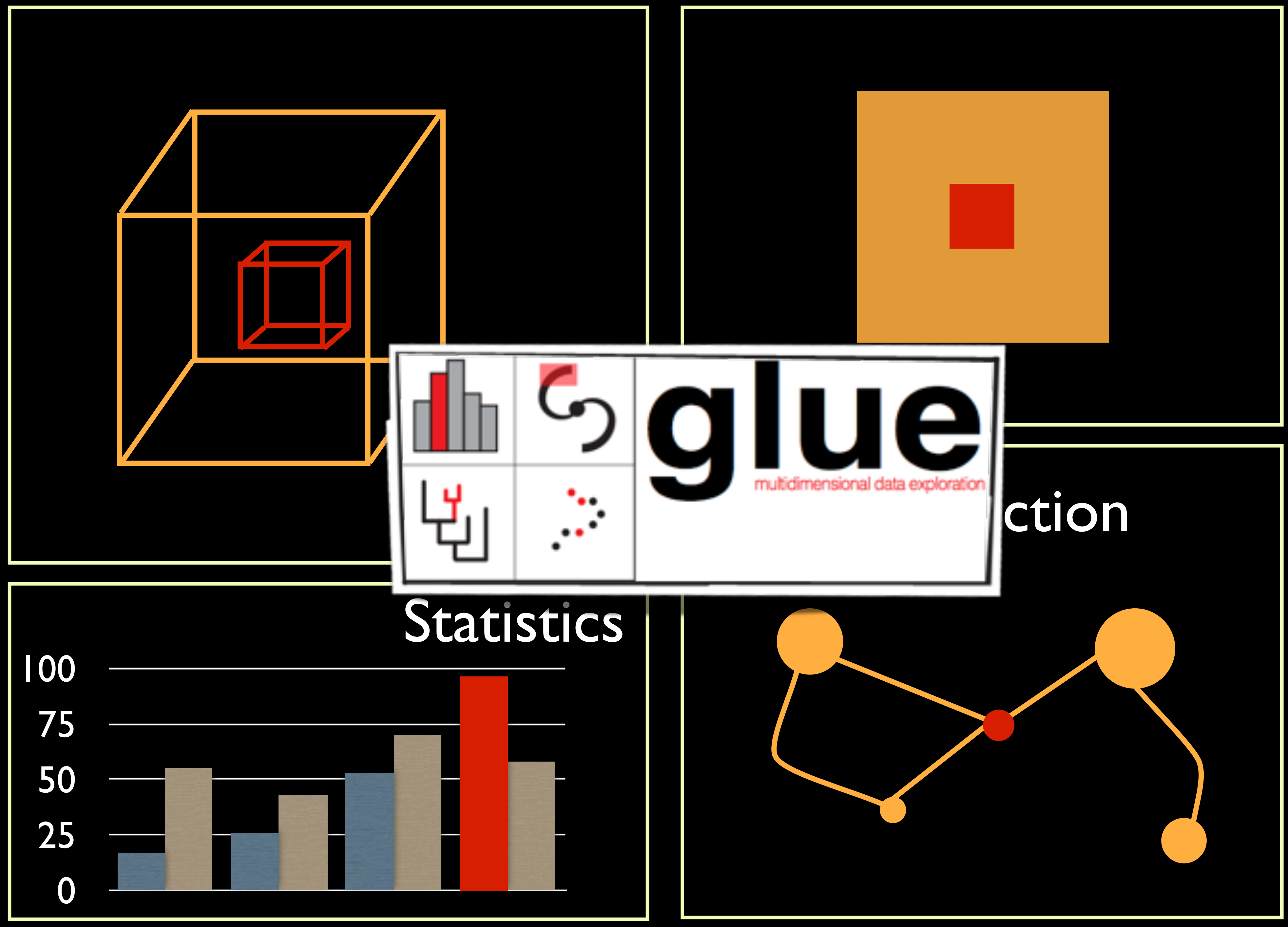


Goodman et al. *Nature*, 2009

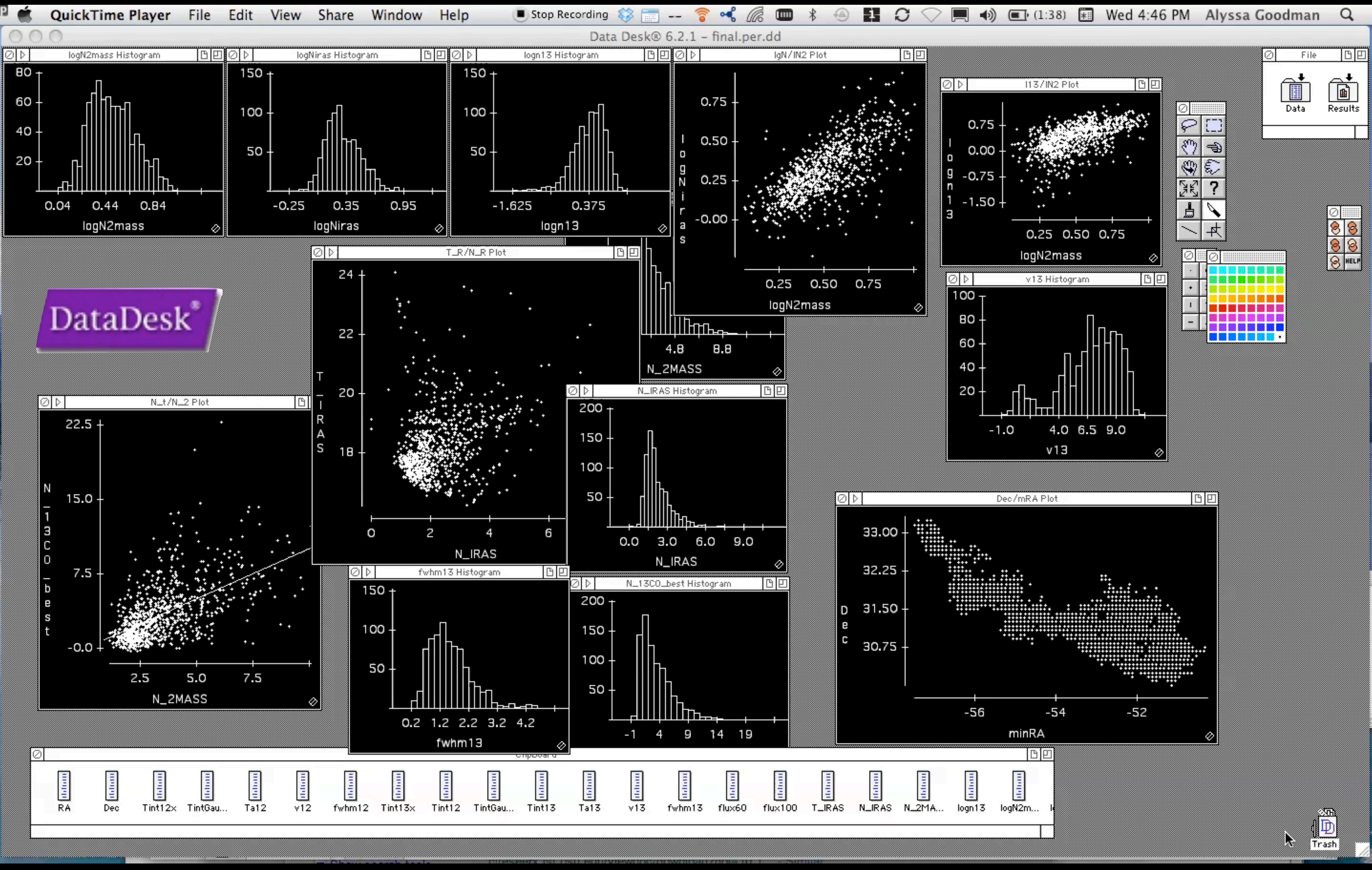


Elde et al. *Nature*, 2008

"Linked Views"



“Linked Views”: DataDesk (est. 1986)



John Tukey's "Four Essentials" (c.1972)

Picturing

Rotation

Isolation

Masking

Selection

and these *"need to work together"*
in a *"dynamic display"*

Brushing

Linking

Results...

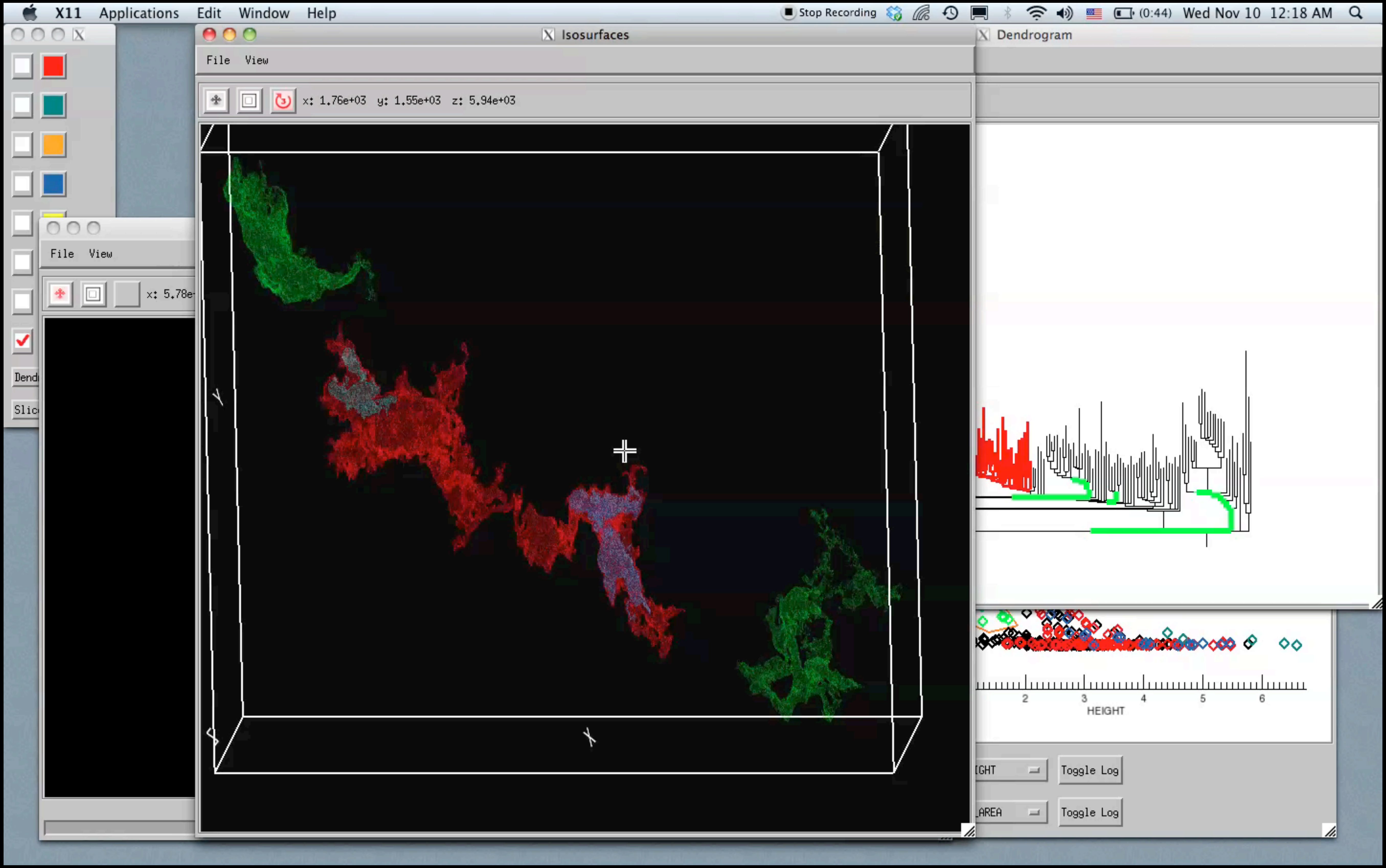
1. for immediate insight
2. as visual source of ideas for statistical algorithms (...relation to SVM)

Warning

"details of control can make or break such a system"

Watch the PRIM-9 video at: http://stat-graphics.org/movies/prim9.html

Exemplar: Linked Dendrogram Views in IDL



Video & implementation: Christopher Beaumont, CfA/UHawaii;
inspired by AstroMed work of Douglas Alan, Michelle Borkin, AG, Michael Halle, Erik Rosolowsky

JOHN TUKEY'S LEGACY

PRIM-9 → DataDesk →

PRIM-H

glue

XGobi → GGobi →

RGGobi

Spotfire →

Polaris → tableau

1970 1980 1990 2000 2010

Astron. Nachr. / AN **333**, No. 5/6, 505 – 514 (2012) / **DOI** 10.1002/asna.201211705

JohnTukey

Principles of high-dimensional data visualization in astronomy

A.A. Goodman*

Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA

Received 2012 May 3, accepted 2012 May 4
Published online 2012 Jun 15

Key words cosmology: large-scale structure – ISM: clouds – methods: data analysis – techniques: image processing – techniques: radial velocities

Astronomical researchers often think of analysis and visualization as separate tasks. In the case of high-dimensional data sets, though, interactive *exploratory data visualization* can give far more insight than an approach where data processing and statistical analysis are followed, rather than accompanied, by visualization. This paper attempts to charts a course toward "linked view" systems, where multiple views of high-dimensional data sets update live as a researcher selects, highlights, or otherwise manipulates, one of several open views. For example, imagine a researcher looking at a 3D volume visualization of simulated or observed data, and simultaneously viewing statistical displays of the data set's properties (such as an x-y plot of temperature vs. velocity, or a histogram of vorticities). Then, imagine that when the researcher selects an interesting group of points in any one of these displays, that the same points become a highlighted subset in all other open displays. Selections can be graphical or algorithmic, and they can be combined, and saved. For tabular (ASCII) data, this kind of analysis has long been possible, even though it has been under-used in astronomy. The bigger issue for astronomy and other "high-dimensional" fields, though, is that no extant system allows for full integration of images and data cubes within a linked-view environment. The paper concludes its history and analysis of the present situation with suggestions that look toward cooperatively-developed open-source modular software as a way to create an evolving, flexible, high-dimensional, linked-view visualization environment useful in astrophysical research.

glue

glue

multidimensional data exploration

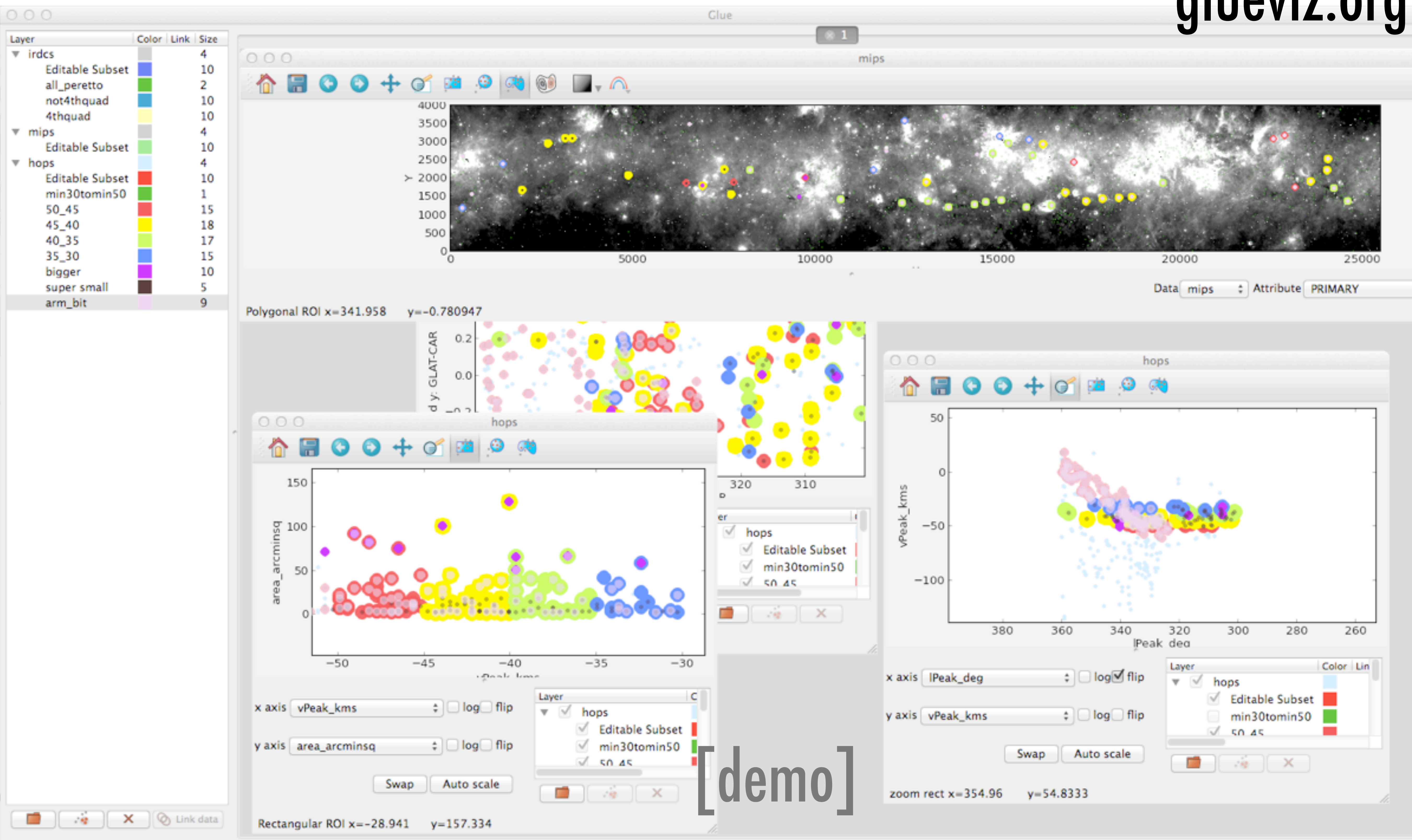
[Beaumont elevator slides]

glueviz.org

glue

multidimensional data exploration

glueviz.org



Glue collaboration: Beaumont, Borkin, Goodman, Pfister, Robitaille

glueing together the Universe

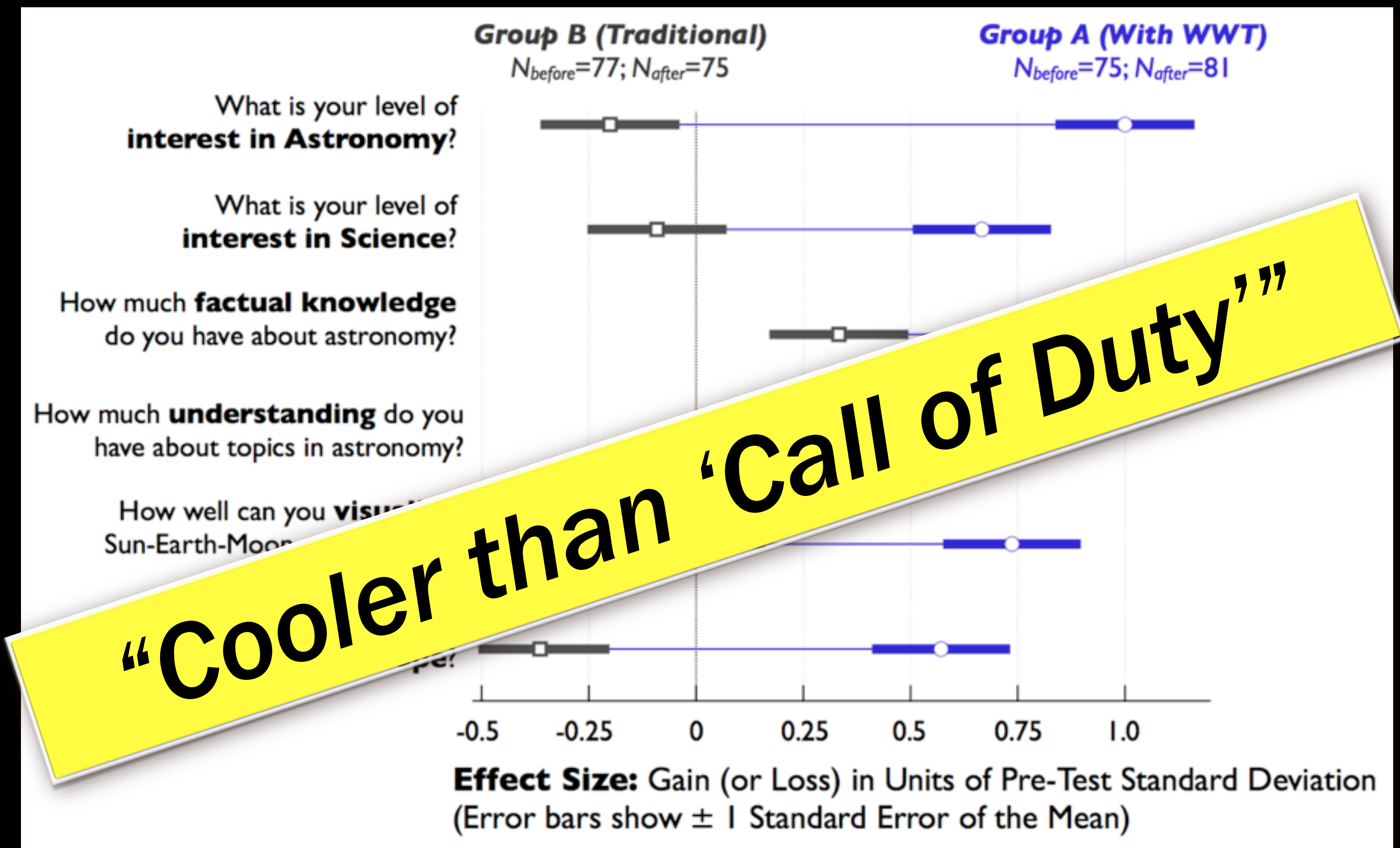
multidimensional data exploration

and the Earth

worldwidetelescope.org glueviz.org

Alyssa A. Goodman • Harvard-Smithsonian Center for Astrophysics • @aagie

WWT Ambassadors

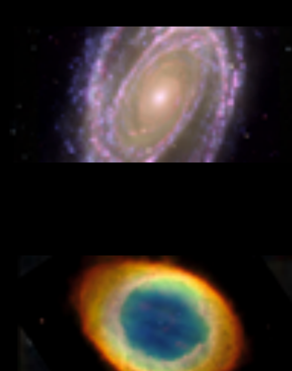


WWT in Higher Ed

edX

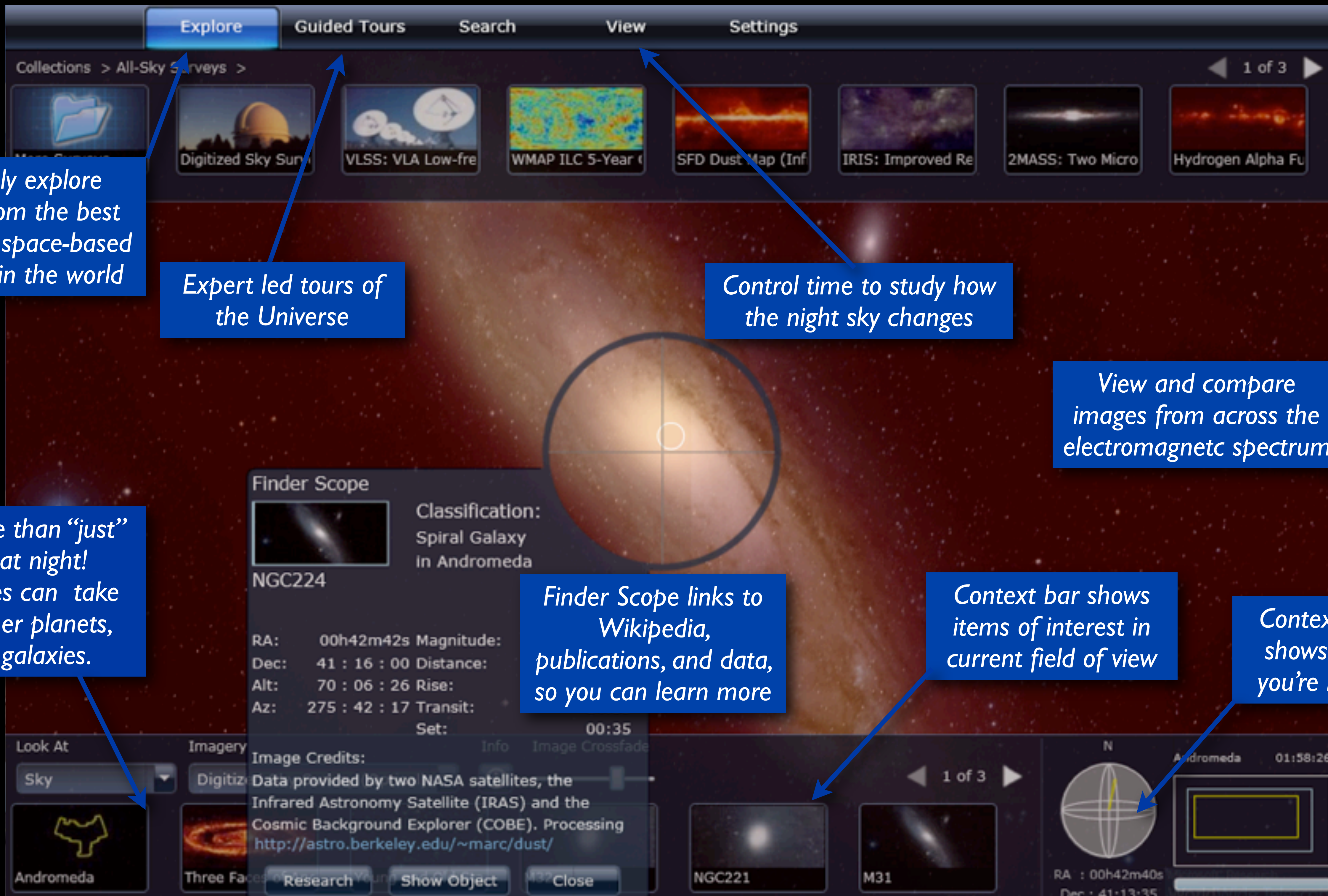
WWT in Research

"Tours"



Microsoft® Research WorldWide Telescope

worldwidetelescope.org



Seamlessly explore imagery from the best ground and space-based telescopes in the world

Expert led tours of the Universe

Control time to study how the night sky changes

View and compare images from across the electromagnetc spectrum

Much more than "just" the sky at night! 3D features can take you to other planets, stars & galaxies.

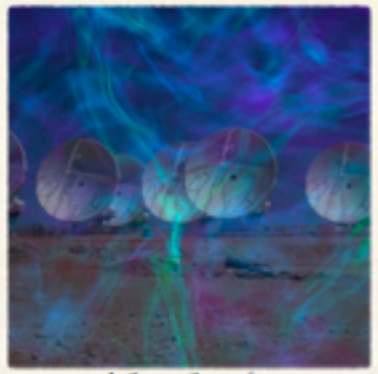
Finder Scope links to Wikipedia, publications, and data, so you can learn more

Context bar shows items of interest in current field of view

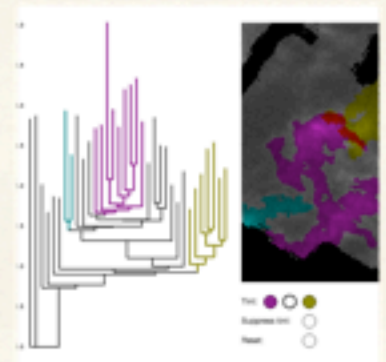
Context globe shows where you're looking.

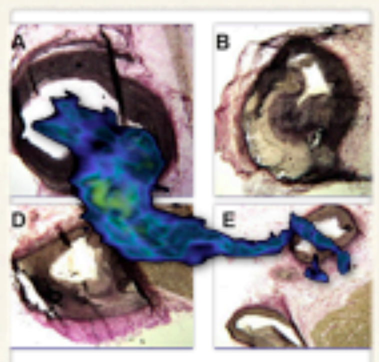
Viz-e-Lab

Projects 2011


"Taste-Testing"


WorldWide Telescope/Ambassadors


Linked Views/Dendrograms



High-D Data Viz


Seamless Astronomy


Wolbach User Experience Lab

ADS Labs

CfA "Astronomy Dataverse"

VAO/Online Astronomy User Group

collaborators/**contacts** at CfA

Seamless Astronomy: Alyssa **Goodman** **Online Astronomy** Group, CfA Data **Archives**: Gus **Muench** **ADS** Group: Alberto **Accomazzi**

WorldWide Telescope Ambassadors: Pat **Udomprasert** High-Dimensional **Data Visualization & Interaction**: Michelle **Borkin**

Wolbach Library Lab at CfA : Christopher **Erdmann** **VAO** at CfA: Pepi **Fabbiano** Social **Networks** in Science: Alberto **Pepe**

Questions about using the Viz-e-Lab? Contact Sarah Block, 5-7331, sblock@cfa.harvard.edu

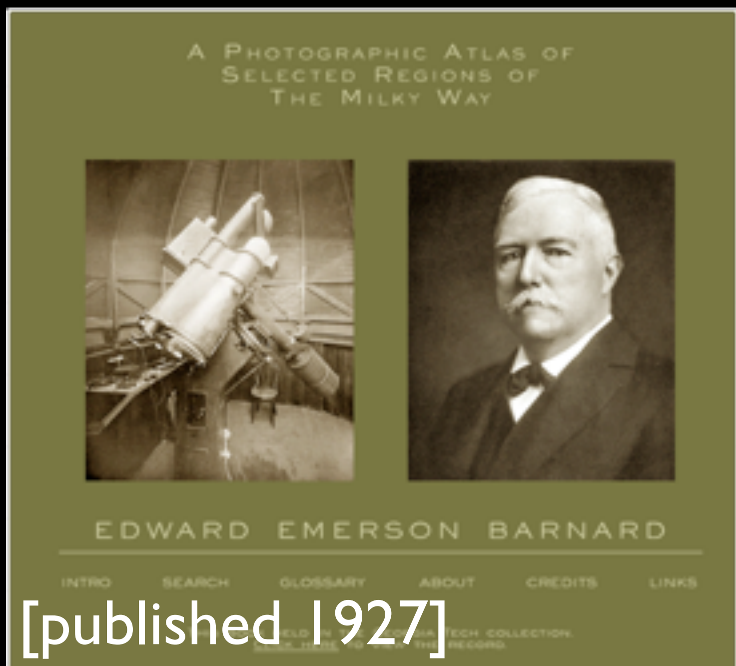
3D Selection

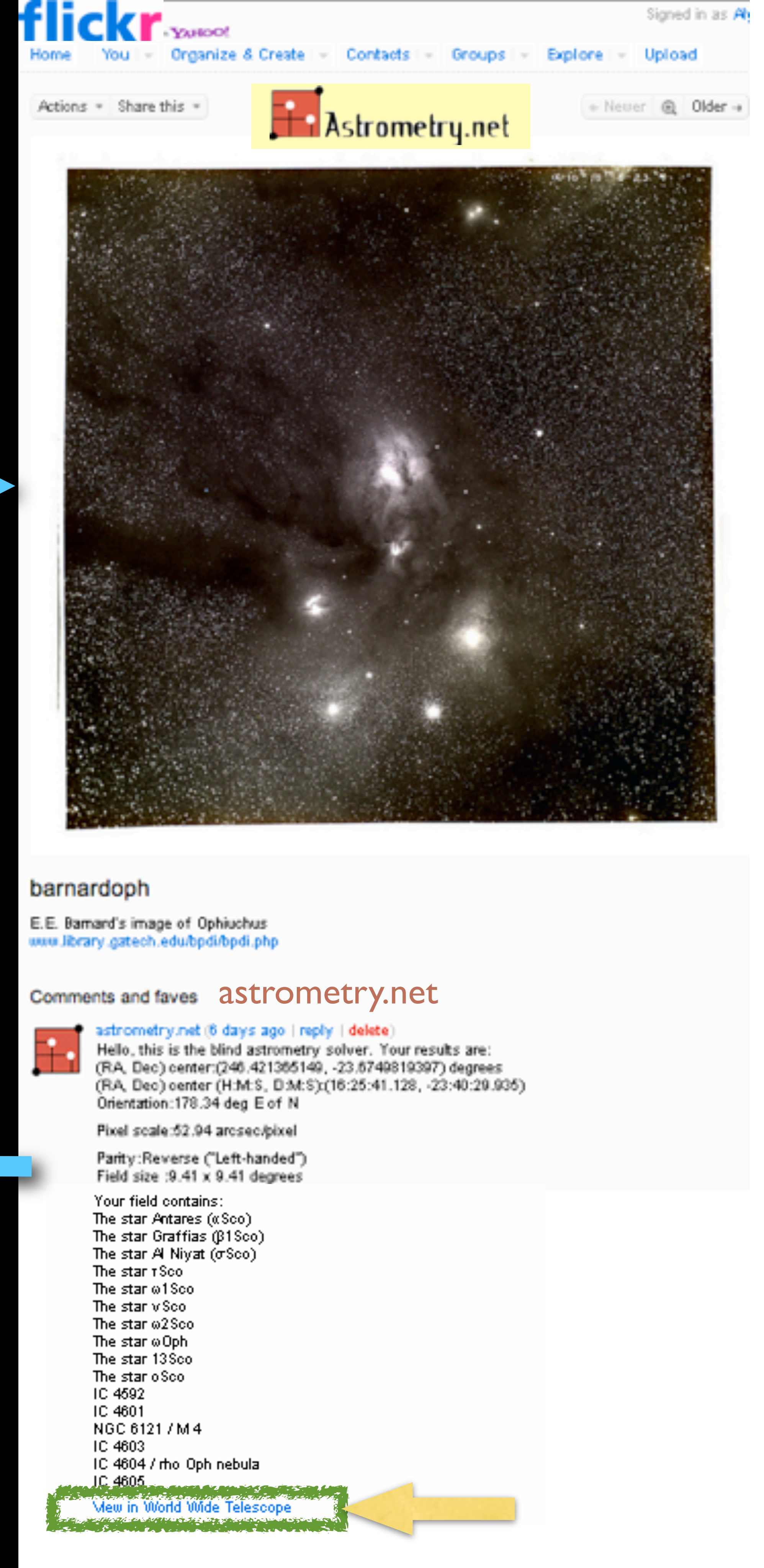
Wh

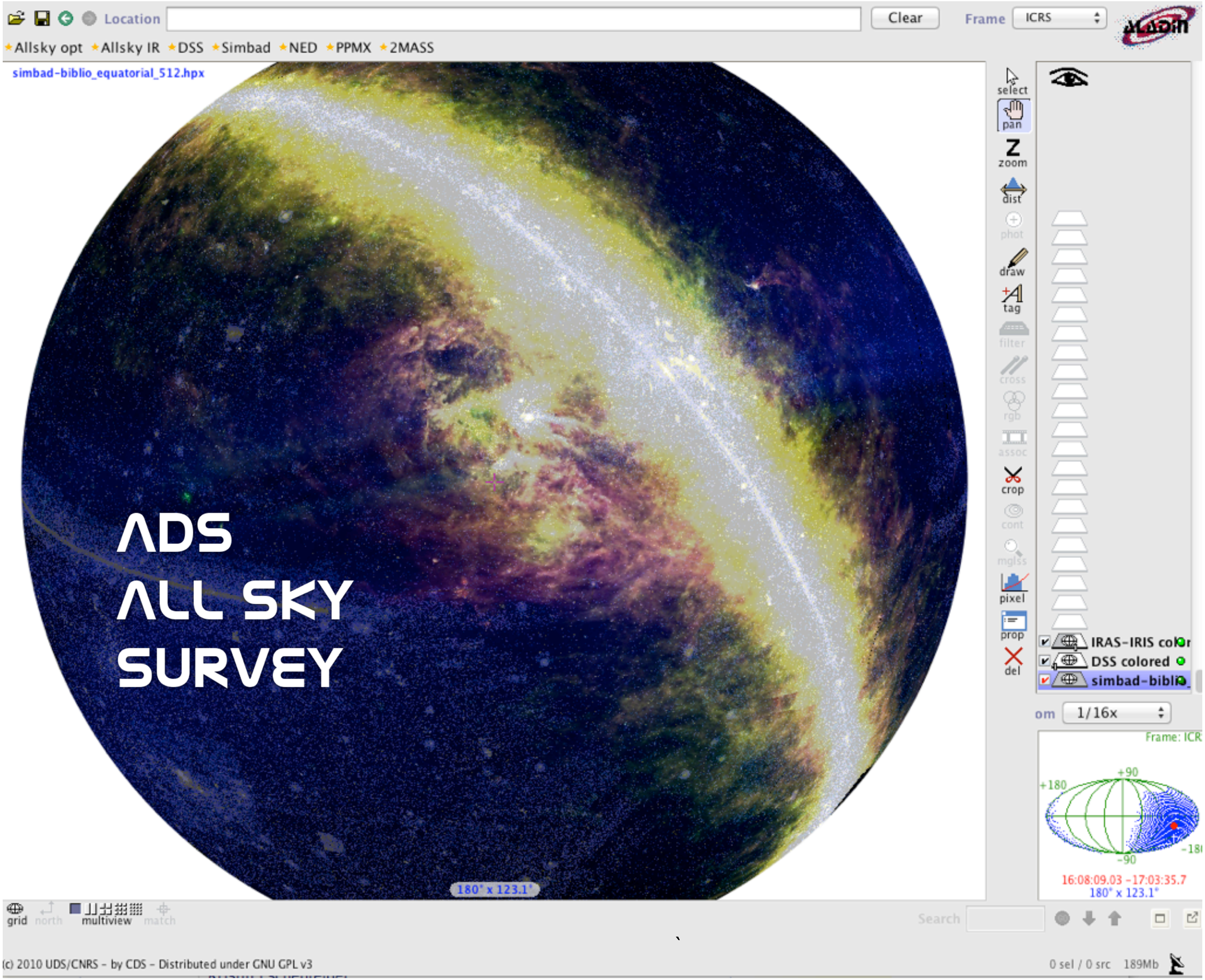
Hov

How … on?

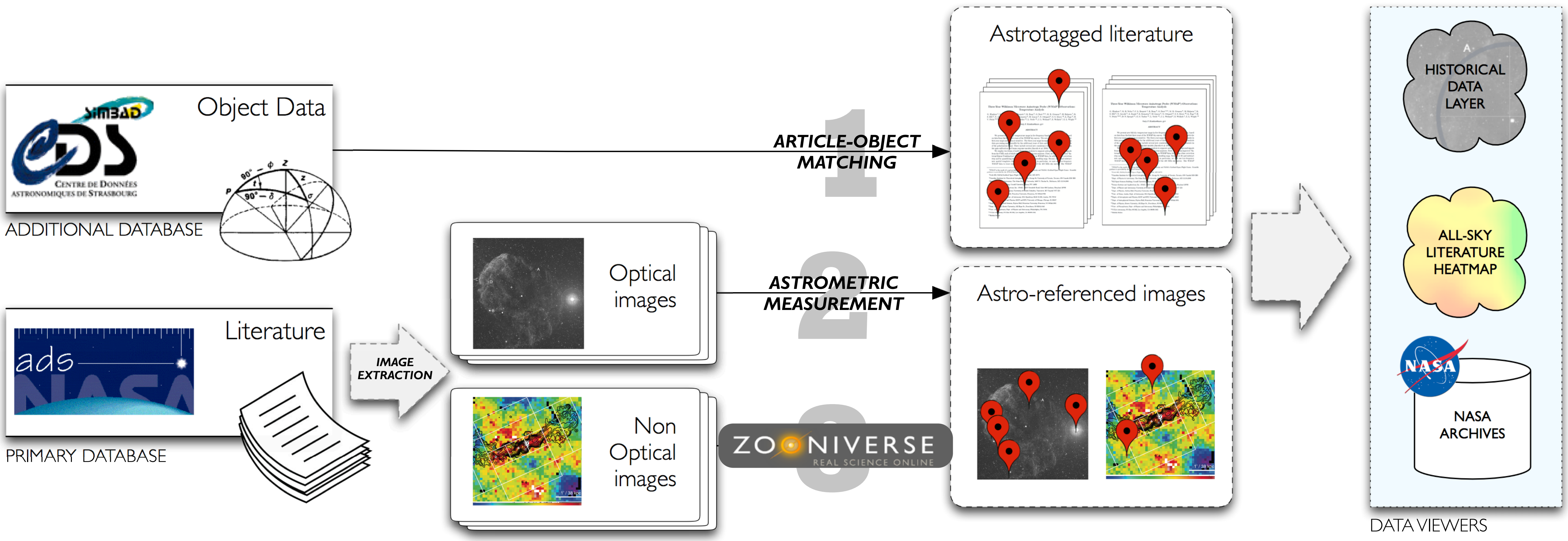
Reviving "Dead" Data






Location
Clear
Frame
ICRS
Allsky opt
Allsky IR
DSS
Simbad
NED
PPMX
2MASS
simbad-biblio_equatorial_512.hpx
ADS
ALL SKY
SURVEY
select
pan
zoom
dist
phot
draw
tag
filter
cross
rgb
assoc
crop
cont
mglss
pixel
prop
del
IRAS-IRIS color
DSS colored
simbad-biblio
1/16x
Frame: ICRS
+90
+180
-90
16:08:09.03 -17:03:35.7
180° x 123.1°
grid
north
multiview
match
Search
(c) 2010 UDS/CNRS - by CDS - Distributed under GNU GPL v3
0 sel / 0 src
189Mb

Seamless Astronomy: ADS All Sky Survey



slide courtesy of Alberto Pepe