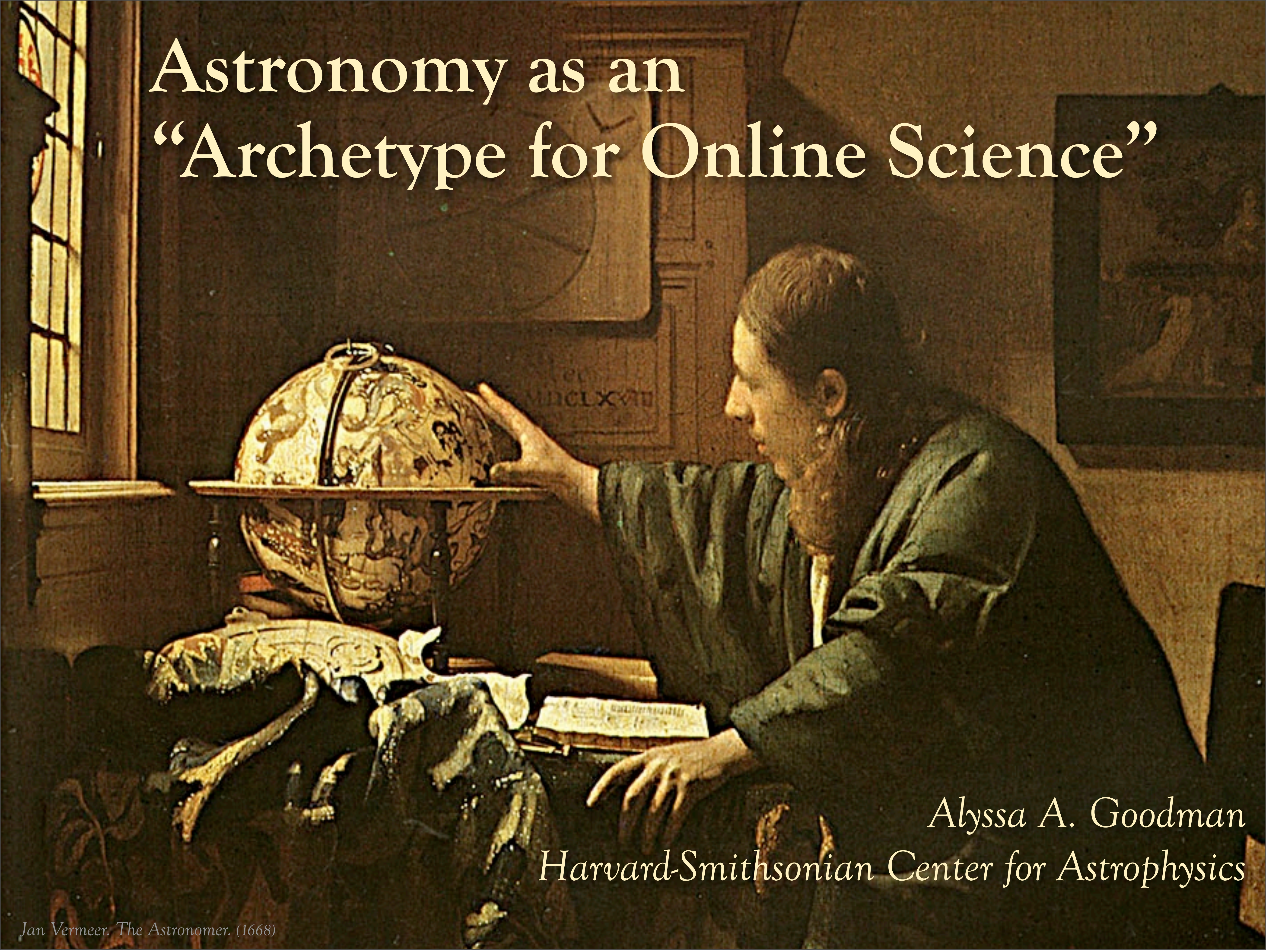
Astronomy as an “Archetype for Online Science”



Alyssa A. Goodman
Harvard-Smithsonian Center for Astrophysics

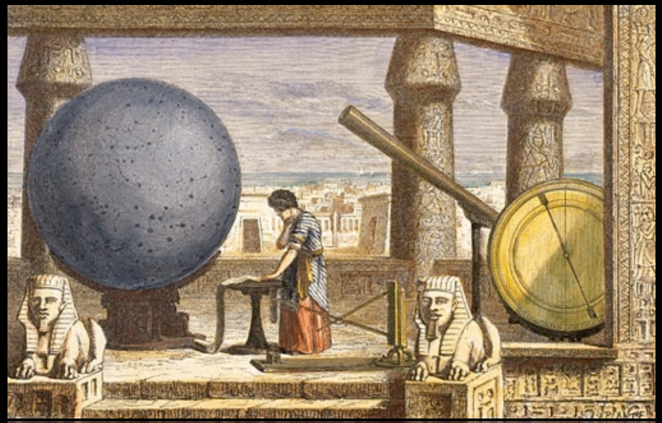
Jan Vermeer. The Astronomer. (1668)

3500 years of Observing

3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD



Observatory Tower, Lincolnshire, UK, c. 1300

3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD

Observatory Tower, Lincolnshire, UK, c. 1300

Galileo, 1600

3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD

Observatory Tower, Lincolnshire, UK, c. 1300

Galileo, 1600

3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD

Observatory Tower, Lincolnshire, UK, c. 1300

Galileo, 1600

The "Scientific Revolution"

3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD

Observatory Tower, Lincolnshire, UK, c. 1300

Galileo, 1600

The "Scientific Revolution"

Reber's Radio Telescope, 1937

3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD

Observatory Tower, Lincolnshire, UK, c. 1300

Galileo, 1600

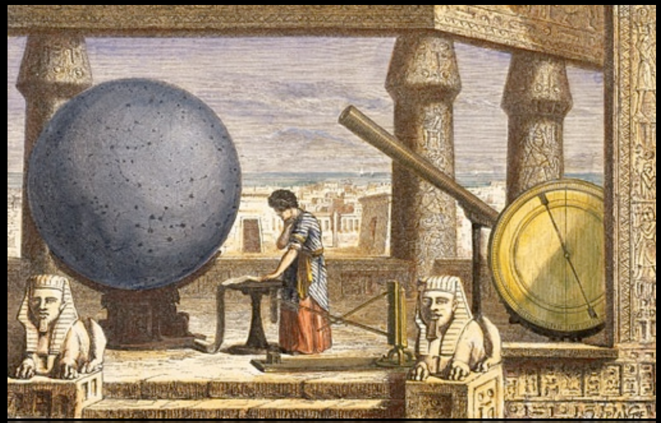
The "Scientific Revolution"

Reber's Radio Telescope, 1937

3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD

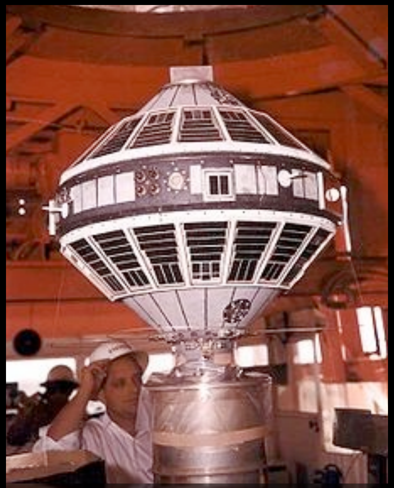


Observatory Tower, Lincolnshire, UK, c. 1300

Galileo, 1600

The "Scientific Revolution"

Reber's Radio Telescope, 1937



NASA/Explorer 7 (Space-based Observing) 1959

3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD

Observatory Tower, Lincolnshire, UK, c. 1300

Galileo, 1600

The "Scientific Revolution"

Reber's Radio Telescope, 1937

NASA/Explorer 7 (Space-based Observing) 1959

"The Internet"

3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD

Observatory Tower, Lincolnshire, UK, c. 1300

Galileo, 1600

The "Scientific Revolution"

Reber's Radio Telescope, 1937

NASA/Explorer 7 (Space-based Observing) 1959

"The Internet"

Long-distance remote-control/ "robotic" telescopes 1990s

3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD

Observatory Tower, Lincolnshire, UK, c. 1300

Galileo, 1600

The "Scientific Revolution"

Reber's Radio Telescope, 1937

NASA/Explorer 7
(Space-based Observing)
1959

"The Internet"

Long-distance remote-control/ "robotic" telescopes
1990s

"Virtual Observatories"
21st century

3500 years of Observing

Stonehenge, 1500 BC

Ptolemy in Alexandria, 100 AD

Observatory Tower, Lincolnshire, UK, c. 1300

Galileo, 1600

The "Scientific Revolution"

Reber's Radio Telescope, 1937

NASA/Explorer 7 (Space-based Observing) 1959

"The Internet"

Long-distance remote-control/ "robotic" telescopes 1990s

"Virtual Observatories" 21st century

2001: a space odyssey

MGM presents a STANLEY KUBRICK PRODUCTION

Super Panavision® and Metrocolor

The World-Wide Telescope, an Archetype for Online Science

Jim Gray
Microsoft Research
Gray@Microsoft.com

Alex Szalay
The Johns Hopkins University
Szalay@jhu.edu

Abstract: Most scientific data will never be directly examined by scientists; rather it will be put into online databases where it will be analyzed and summarized by computer programs. Scientists increasingly see their instruments through online scientific archives and analysis tools, rather than examining the raw data. Today this analysis is primarily driven by scientists asking queries, but scientific archives are becoming active databases that self-organize and recognize interesting and anomalous facts as data arrives. In some fields, data from many different archives can be cross-correlated to produce new insights. Astronomy presents an excellent example of these trends; and, federating Astronomy archives presents interesting challenges for computer scientists.

Introduction

Computational Science is a new branch of most disciplines. A thousand years ago, science was primarily *empirical*. Over the last 500 years each discipline has grown a *theoretical* component. Theoretical models often motivate experiments and generalize our understanding. Today most disciplines have both empirical and theoretical branches. In the last 50 years, most disciplines have grown a third, *computational* branch (e.g. empirical, theoretical, and computational ecology, or physics, or linguistics.)

statistics among sets of data points in a metric space. Pair-algorithms on N points scale as N^2. If the data increase a thousand fold, the work and time can grow by a factor of a million. Many clustering algorithms scale even worse. These algorithms are infeasible for terabyte-scale datasets.

The new online science needs new data mining algorithms that use near-linear processing, storage, and bandwidth, and that can be executed in parallel. Unlike current algorithms that give exact answers, these algorithms will likely be heuristic and give approximate answers [Connolly, Szapudi].

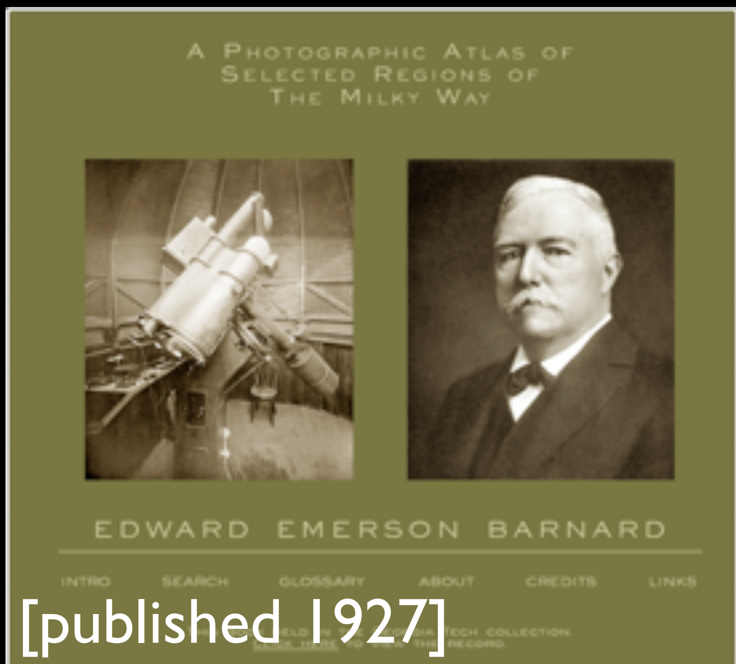
Astronomy as an Archetype for Online Science

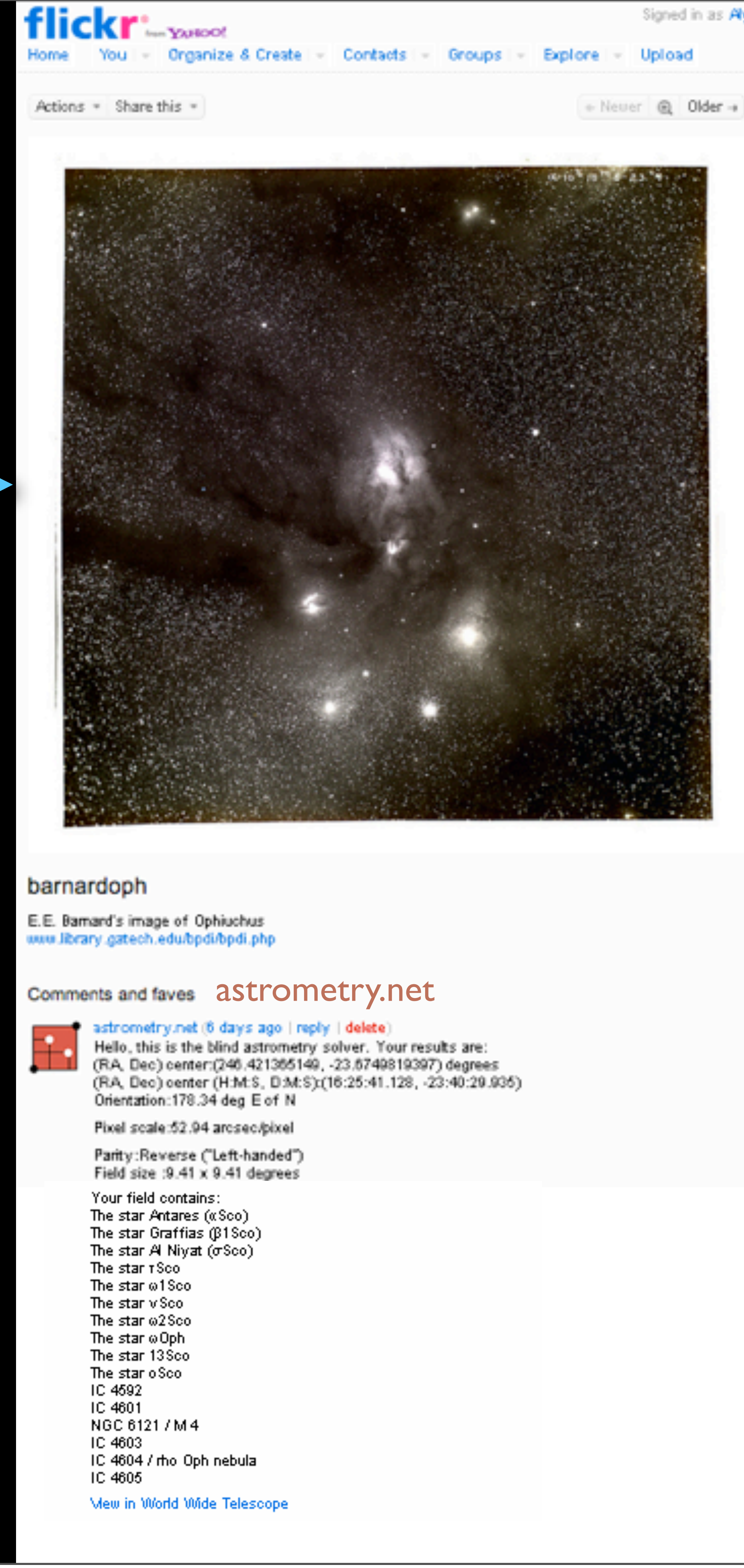
Astronomy exemplifies these phenomena. For thousands of years astronomy was primary empirical with few theoretical models. Theoretical astronomy began with Kepler is now co-equal with observation. Astronomy was early to adopt computational techniques to model stellar and galactic formation and celestial mechanics. Today, simulation is an important part of the field – producing new science, and solidifying our grasp of existing theories.

Astronomers are building telescopes that produce terabytes of data each year -- soon terabytes per night. In the old

Hidden Metadata

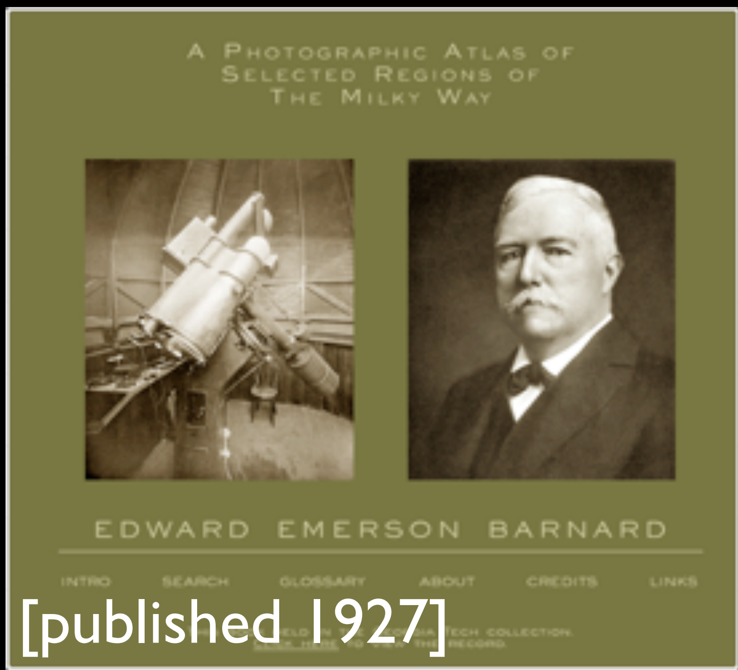
astrometry.net + flickr + WWT

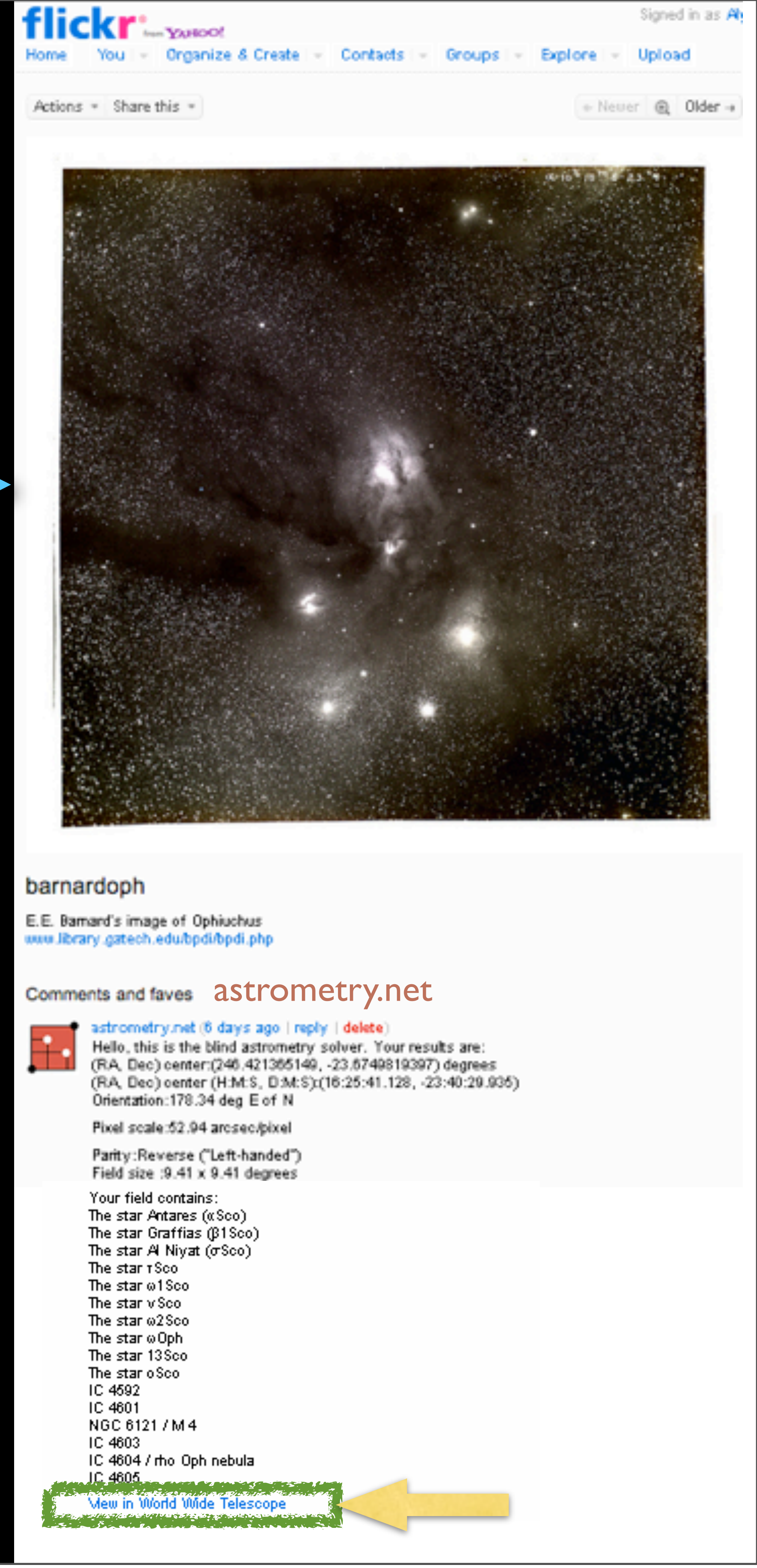




Hidden Metadata

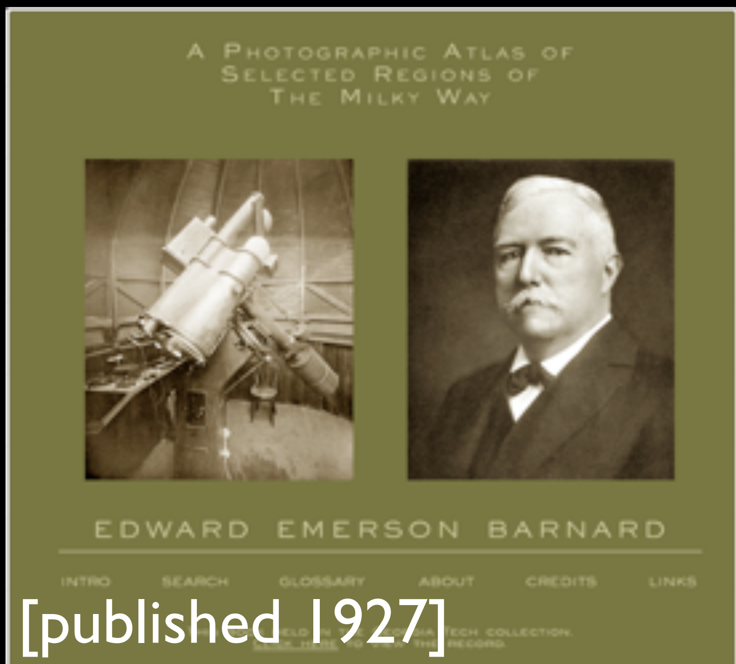
astrometry.net + flickr + WWT



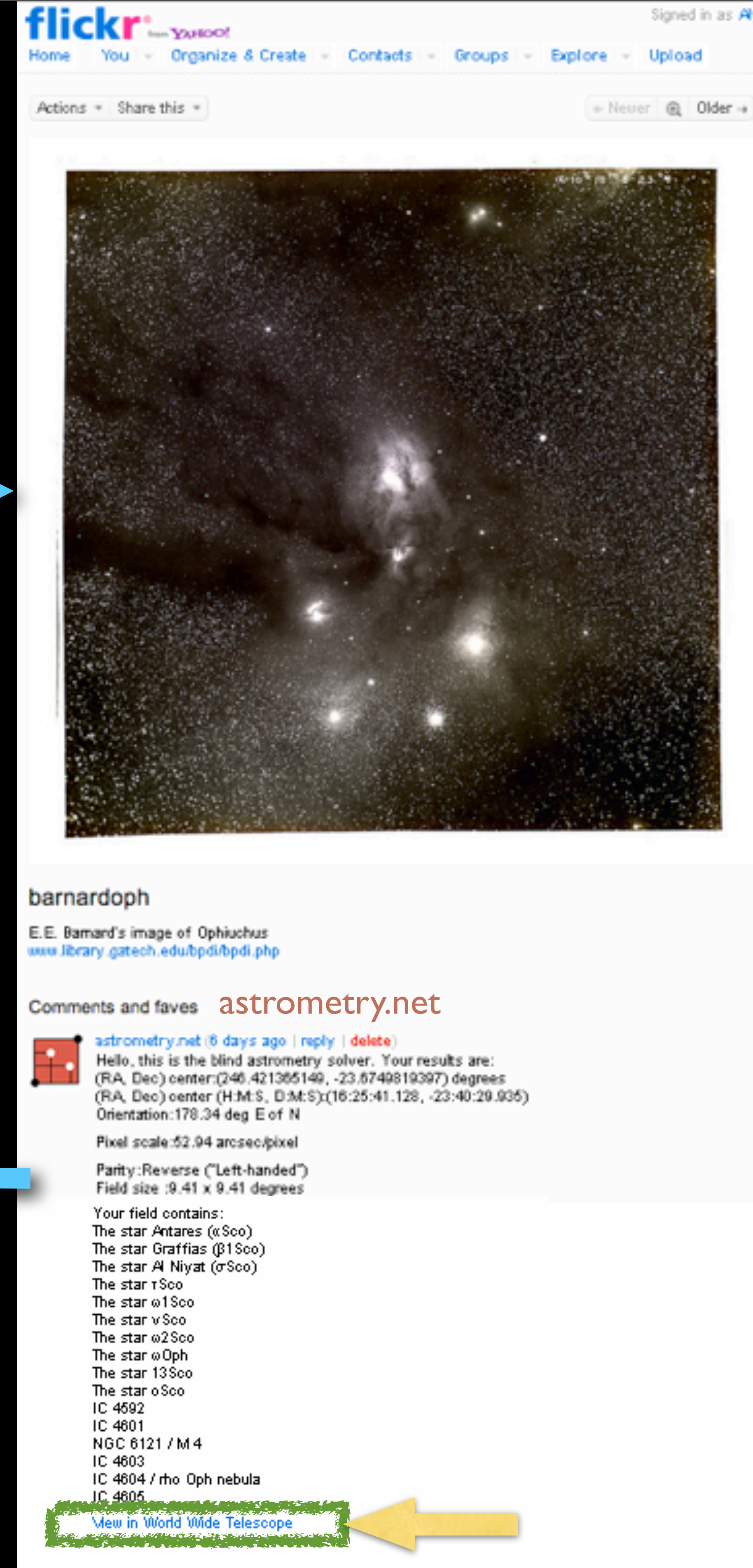


Hidden Metadata

astrometry.net + flickr + WWT



[published 1927]



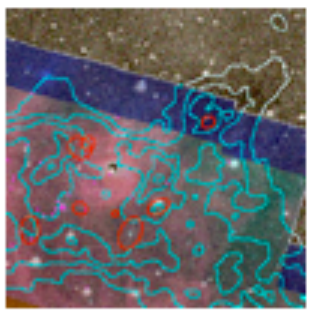
astrometry.net

Experience WWT at **worldwidetelescope.org**

AG & "Online Science"

Research

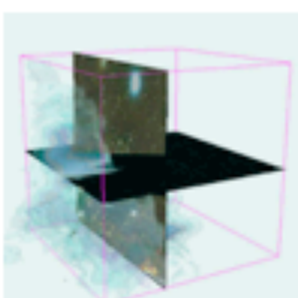
In my *Astronomy* research, I am primarily interested in how the gas in galaxies constantly re-arranges itself over huge time spans to constantly form new stars. I have also had a long-standing interest in data *visualization*, and in improving the use of *computers* in all aspects of scientific research. I teach a course at Harvard called "The Art of Numbers," and I am very involved in the WorldWide Telescope Project, which brings astronomical data to everyone through an interface that demonstrates data delivery for the 21st Century of "e-Science."



COMPLETE
The COordinated Molecular Probe Line Extinction Thermal Emission Survey of Star Forming Regions

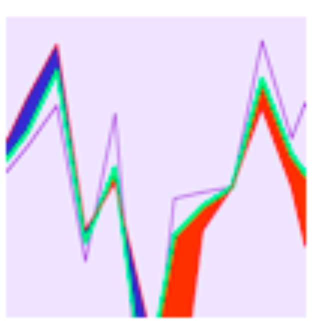
Star Formation Taste Tests
A community of theorists, numericists, and observers working together to compare "observed" simulations with the real Universe.

To sign up email rshetty AT cfa.harvard.edu

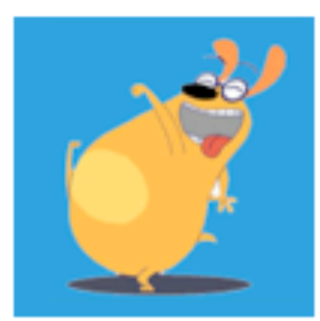


Astronomical Medicine
Exploiting the intersection of astronomical and medical image display and analysis needs to accelerate insight in both fields.

WorldWide Telescope
A beautiful portal to all of Astronomy, for astronomers of all ages and skill levels
http://www.cfa.harvard.edu/WWTAmbassadors



Visualization
Improving the communication of science through imagery and animation



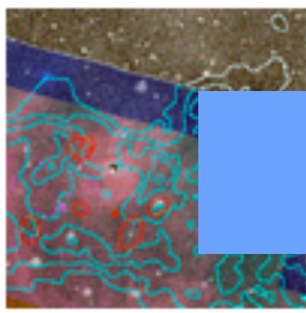
Science for Everyone
Enhancing the public understanding of science

http://www.cfa.harvard.edu/~agoodman/

AG & "Online Science"

Research

In my *Astronomy* research, I am primarily interested in how the gas in galaxies constantly re-arranges itself over huge time spans to constantly form new stars. I have also had a long-standing interest in data *visualization*, and in improving the use of *computers* in all aspects of scientific research. I teach a course at Harvard called "The Art of Numbers," and I am very involved in the WorldWide Telescope Project, which brings astronomical data to everyone through an interface that demonstrates data delivery for the 21st Century of "e-Science."

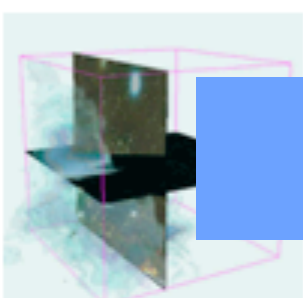


COMPLETE
The COordinated Molecular Probe

Data

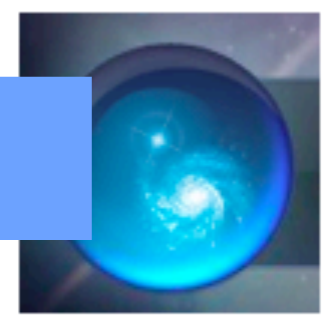
Star Formation Taste Tests
A community of theorists, numericists, and

Simulation

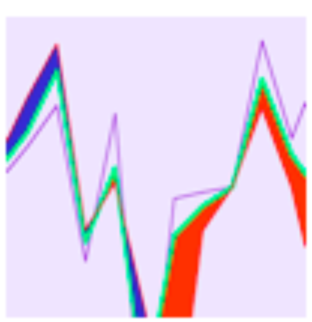


Astronomical Medicine
Exploiting the intersection of

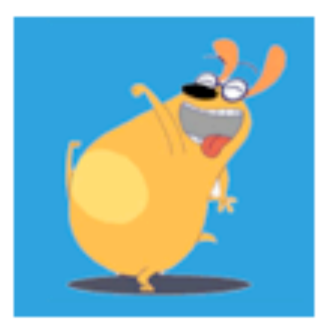
Publishing



WorldWide Telescope
A beautiful portal to all of Astronomy, for astronomers of all ages and skill levels
http://www.cfa.harvard.edu/WWTAmbassadors



Visualization
Improving the communication of science through imagery and animation



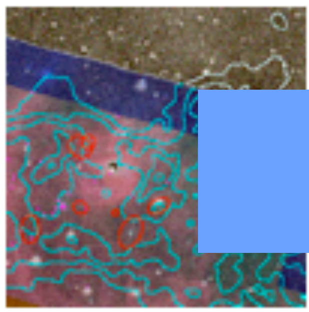
Science for Everyone
Enhancing the public understanding of science

http://www.cfa.harvard.edu/~agoodman/

AG & "Online Science"

Research

In my *Astronomy* research, I am primarily interested in how the gas in galaxies constantly re-arranges itself over huge time spans to constantly form new stars. I have also had a long-standing interest in data *visualization*, and in improving the use of *computers* in all aspects of scientific research. I teach a course at Harvard called "The Art of Numbers," and I am very involved in the WorldWide Telescope Project, which brings astronomical data to everyone through an interface that demonstrates data delivery for the 21st Century of "e-Science."



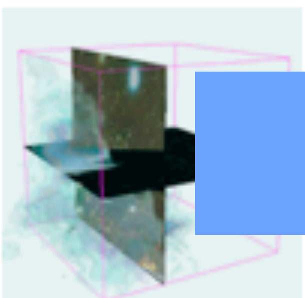
COMPLETE
The COordinated Molecular Probe

Data

COMPLETE

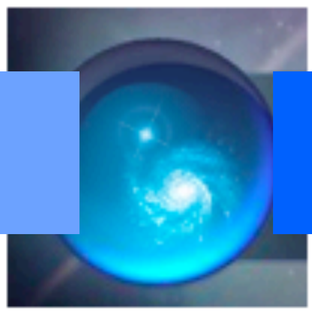
Star Formation Taste Tests
A community of theorists, numericists, and

Simulation



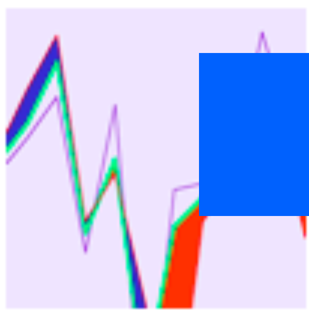
Astronomical Medicine

Publishing



WorldWide Telescope

e-Science Tools



Visualization

Viz

Science for Everyone

Outreach

http://www.cfa.harvard.edu/~agoodman/

Publishing

Data | Simulation

e-Science Tools | Viz | Outreach

Publishing

Data | Simulation

e-Science Tools | Viz | Outreach

Publishing

Data

Simulation

e-Science Tools

Viz

Outreach

Publishing

Data

Simulation

WorldWide Telescope

e-Science Tools

Viz

Outreach

Publishing

Data | Simulation

WorldWide Telescope Ambassadors

e-Science Tools | Viz | Outreach

Publishing

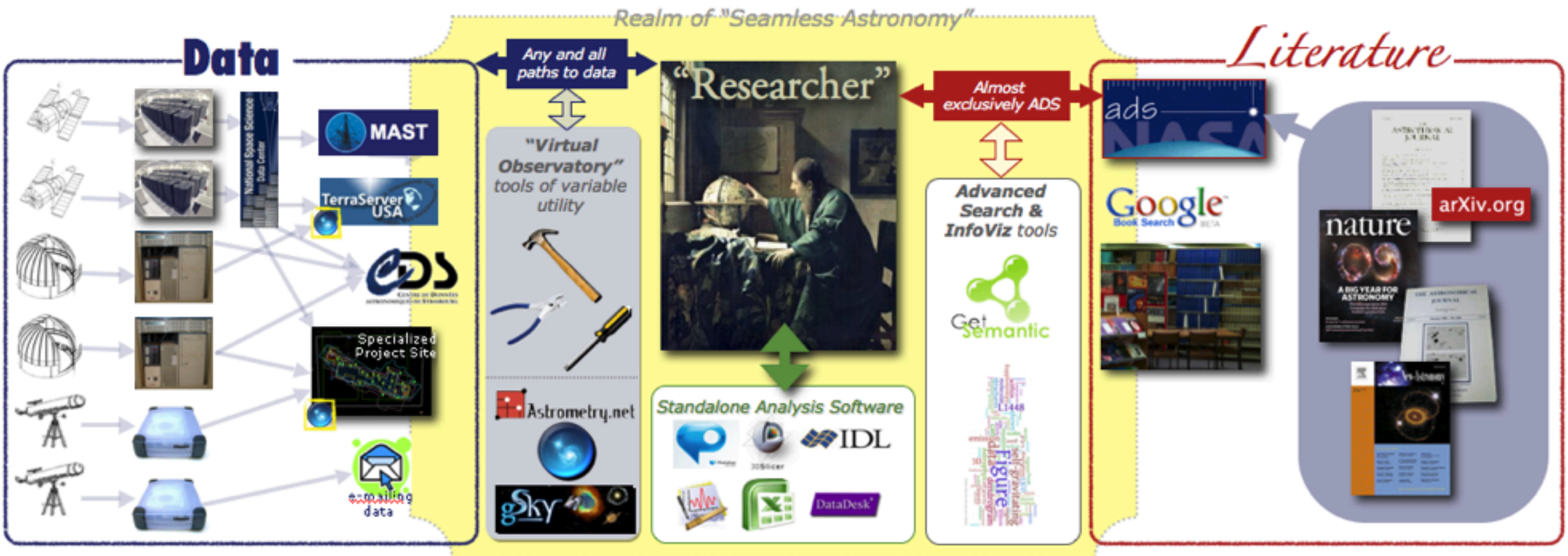
Data | Simulation

VAO | 3D PDF | WorldWide Telescope | Ambassadors

e-Science Tools | Viz | Outreach

SEAMLESS ASTRONOMY

Linking scientific data, publications, and communities



SEAMLESS ASTRONOMY

Linking scientific data, publications, and communities

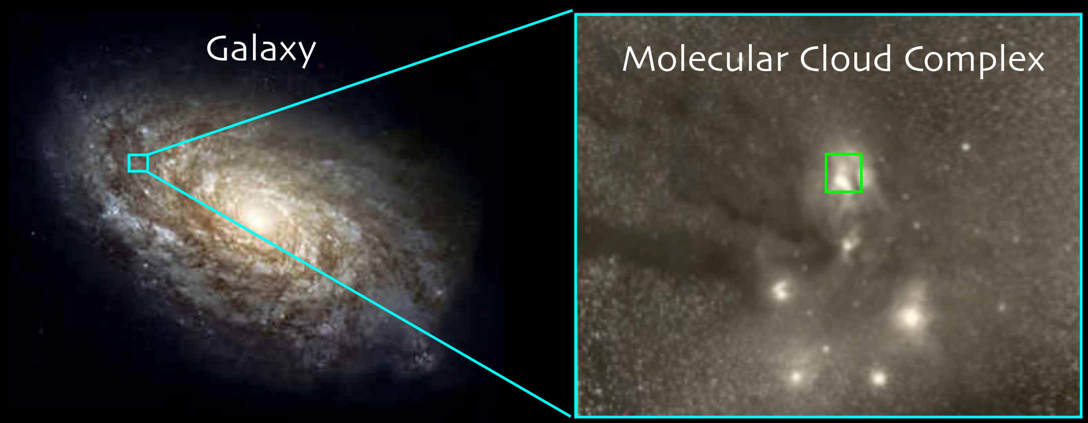
Why?

Star Formation

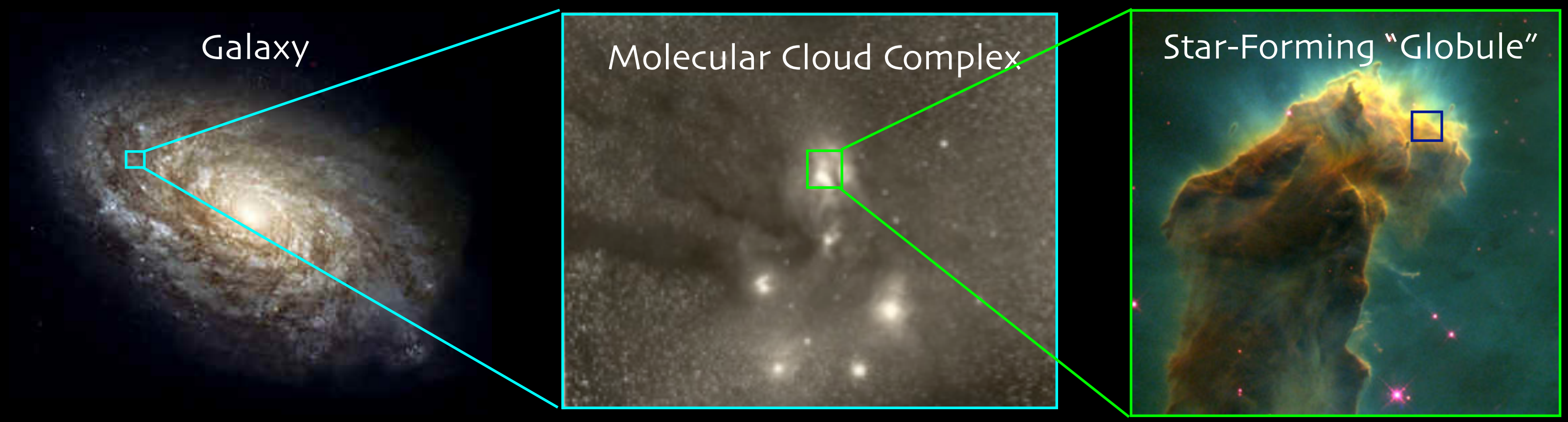
Star (and Planet, and Moon) Formation 101

Star (and Planet, and Moon) Formation 101

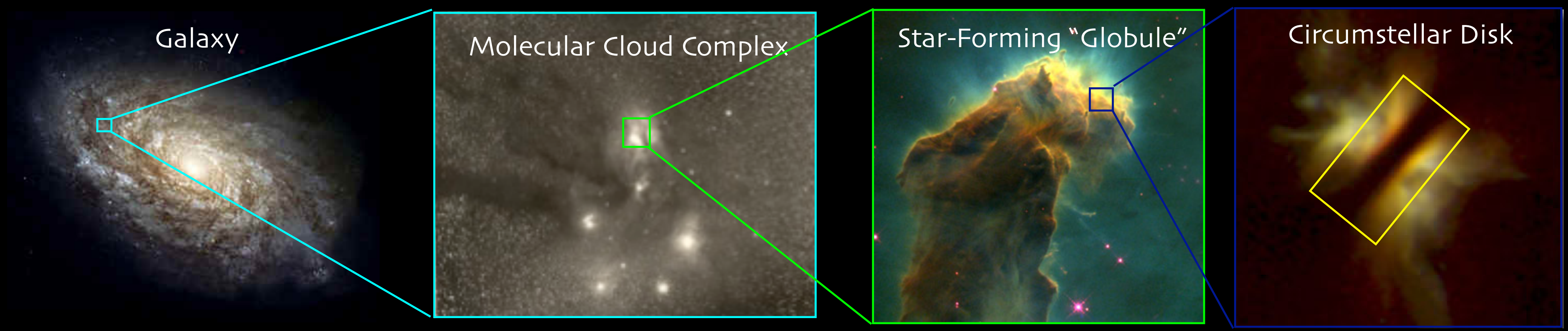
Star (and Planet, and Moon) Formation 101



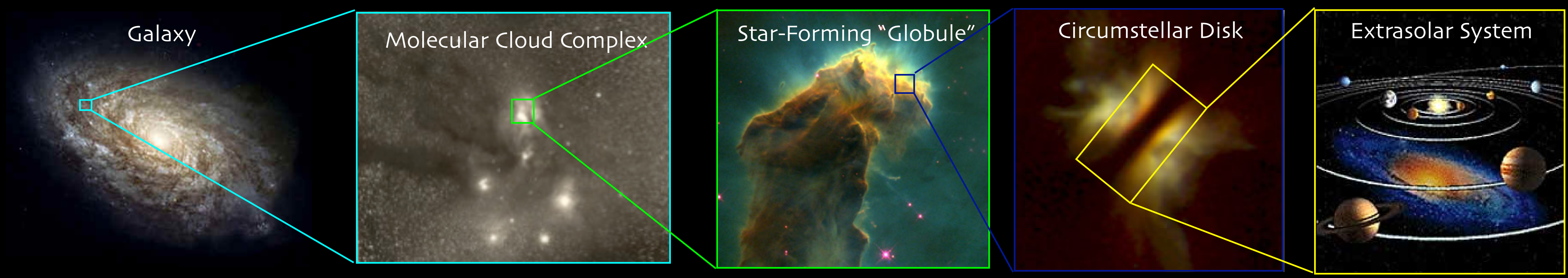
Star (and Planet, and Moon) Formation 101



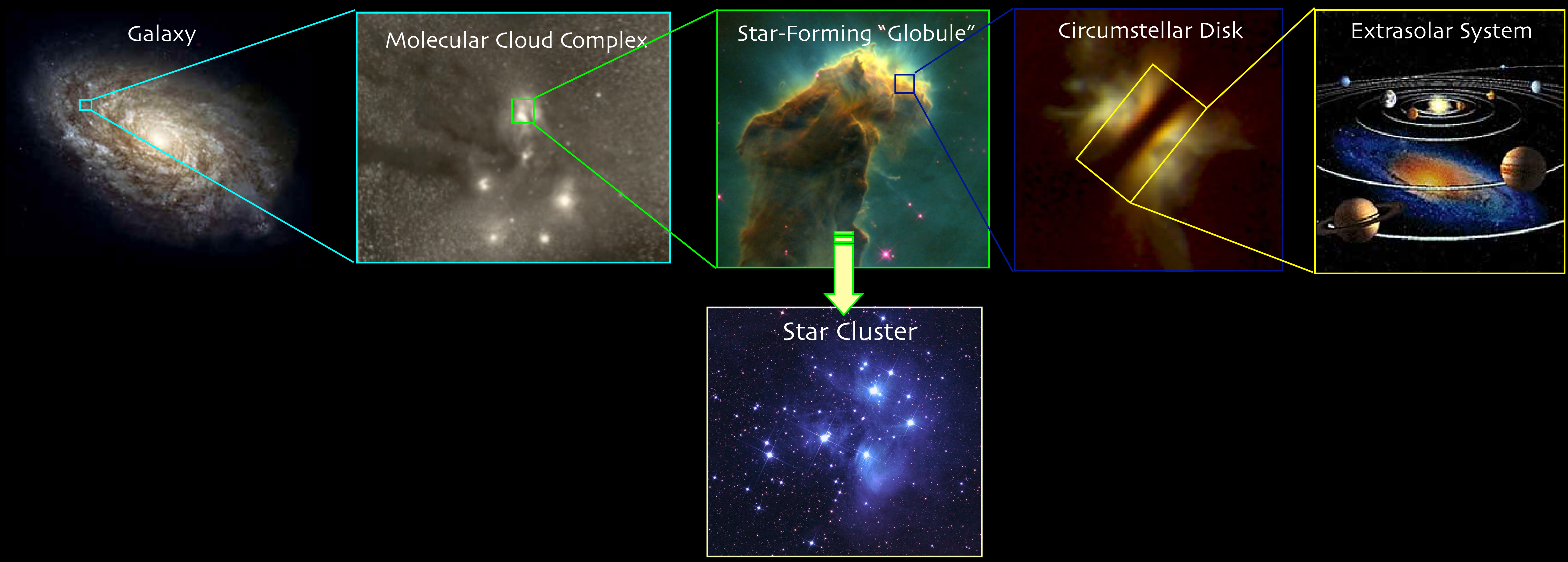
Star (and Planet, and Moon) Formation 101



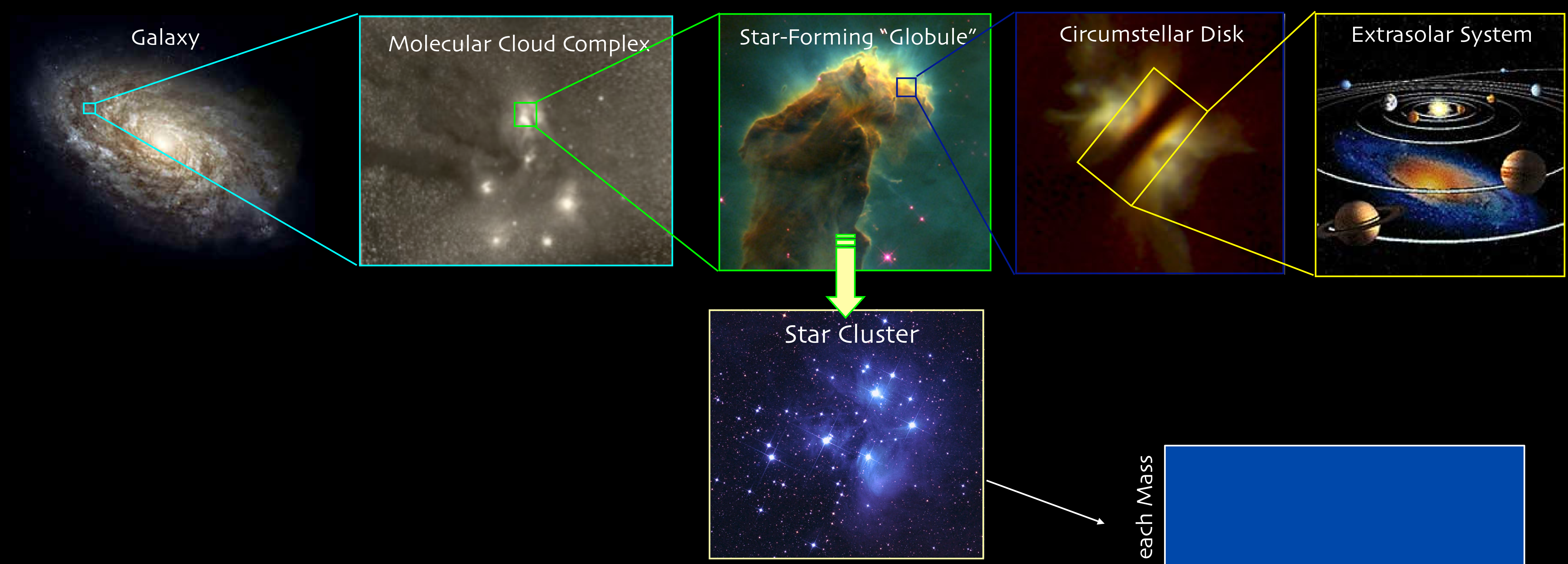
Star (and Planet, and Moon) Formation 101

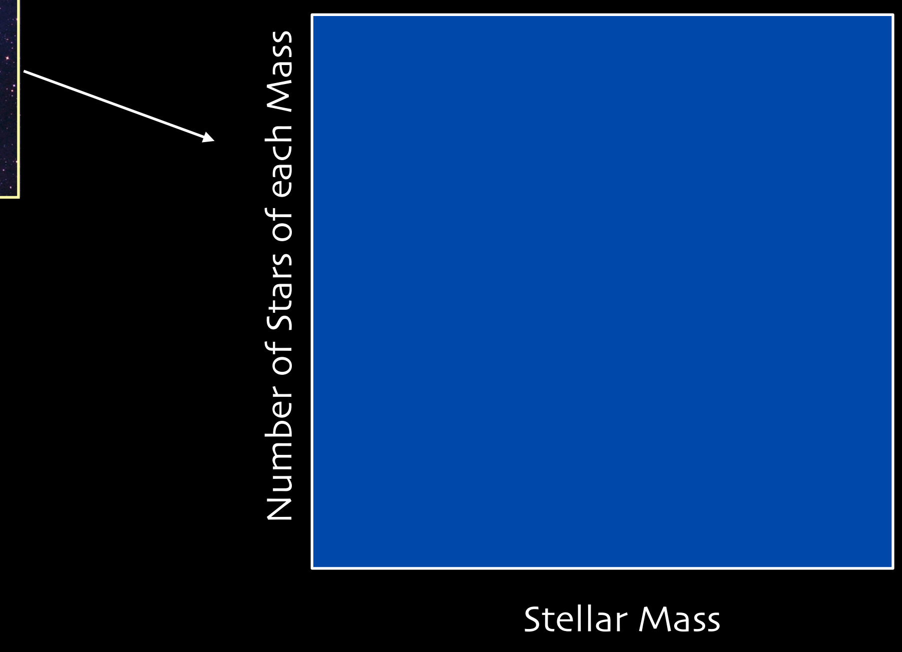


Star (and Planet, and Moon) Formation 101

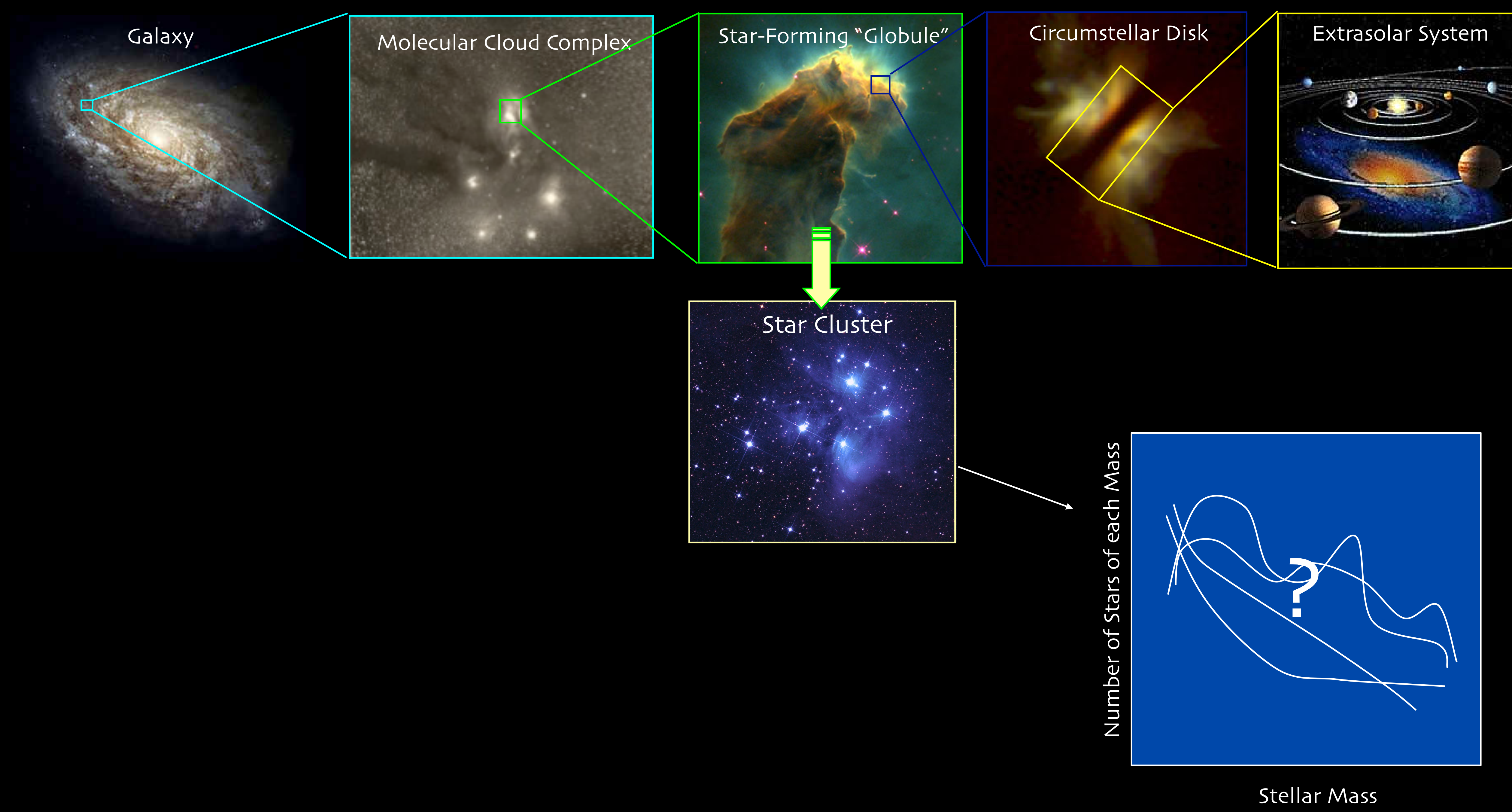


Star (and Planet, and Moon) Formation 101

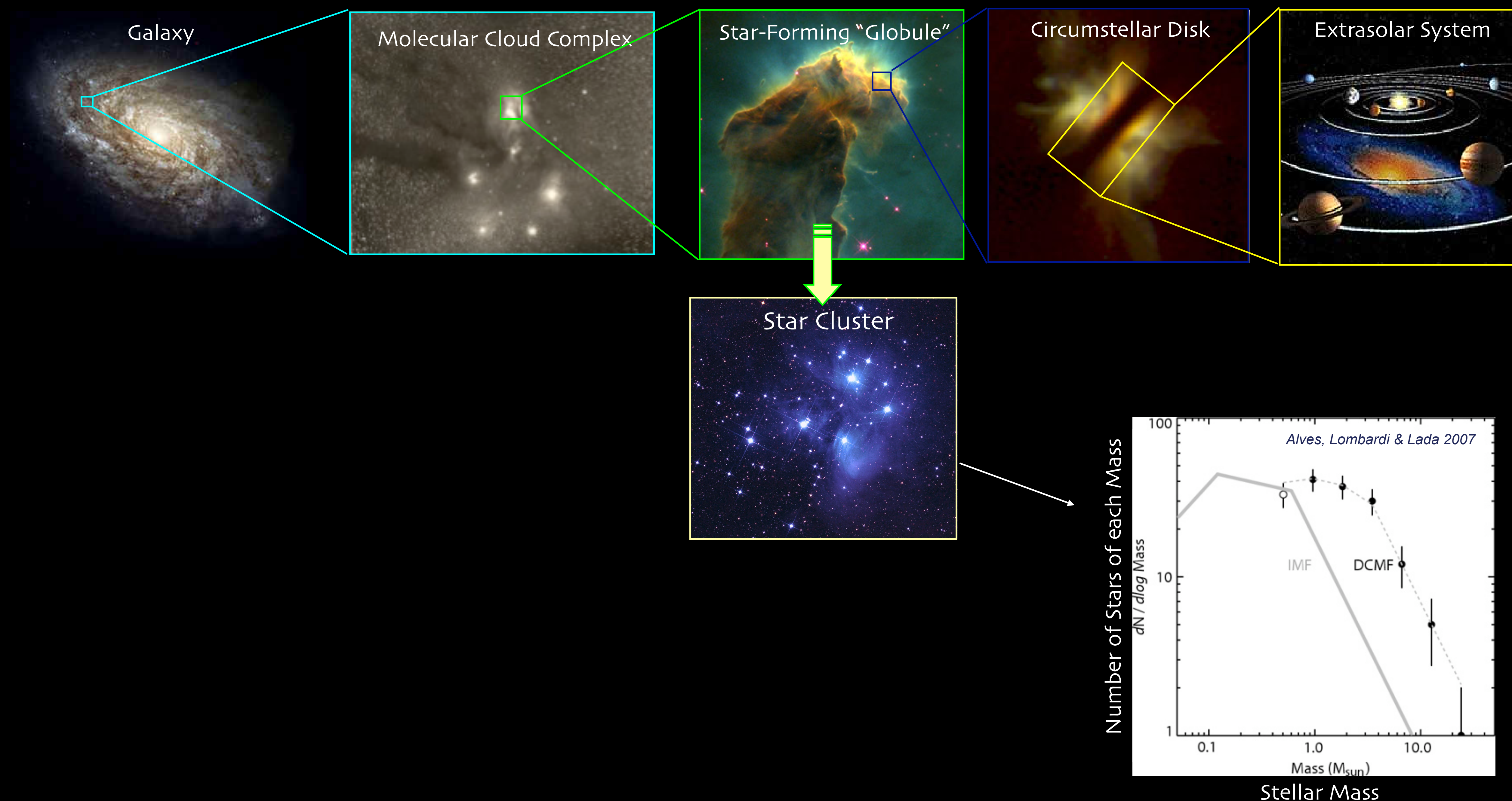




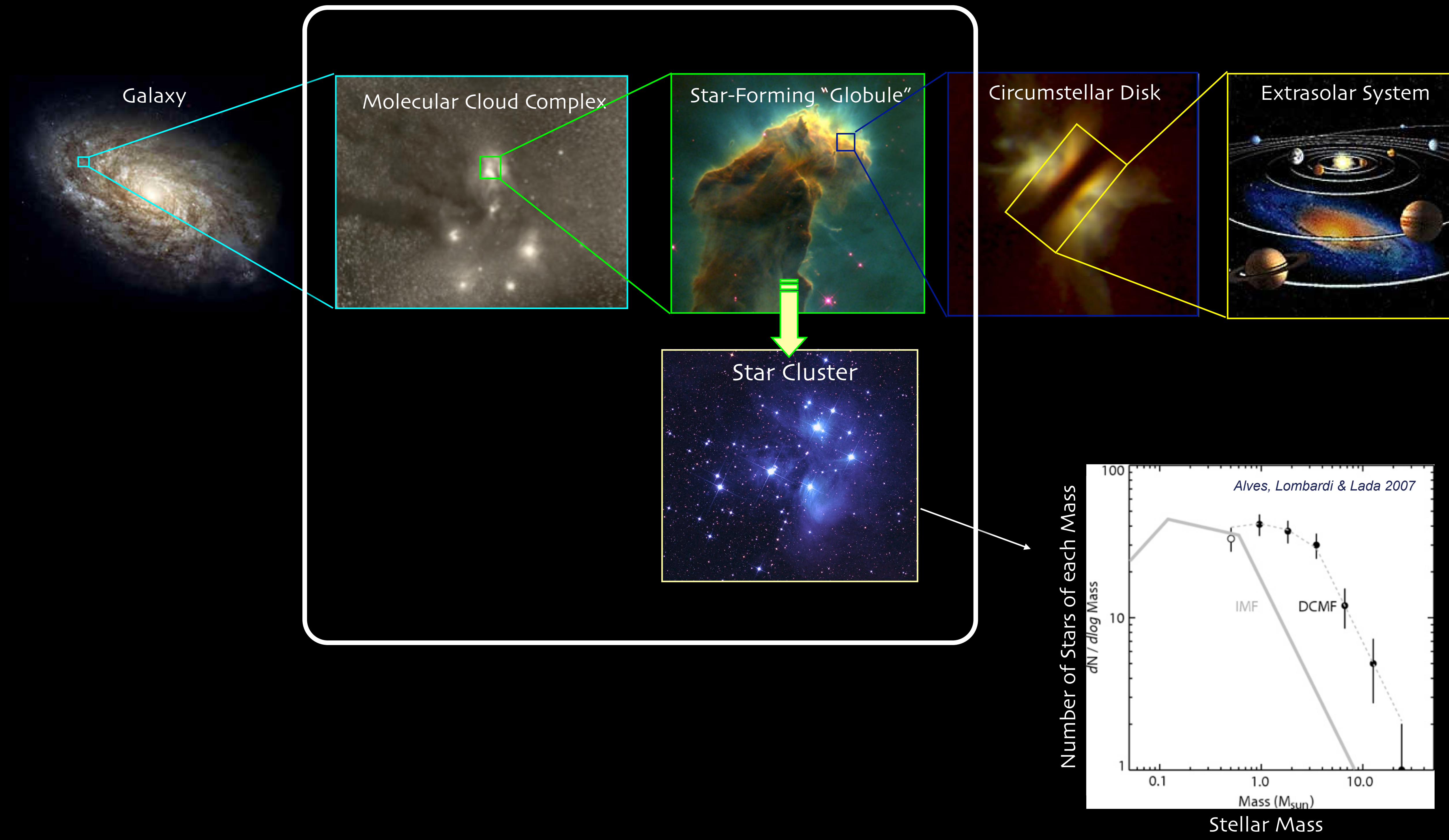
Star (and Planet, and Moon) Formation 101

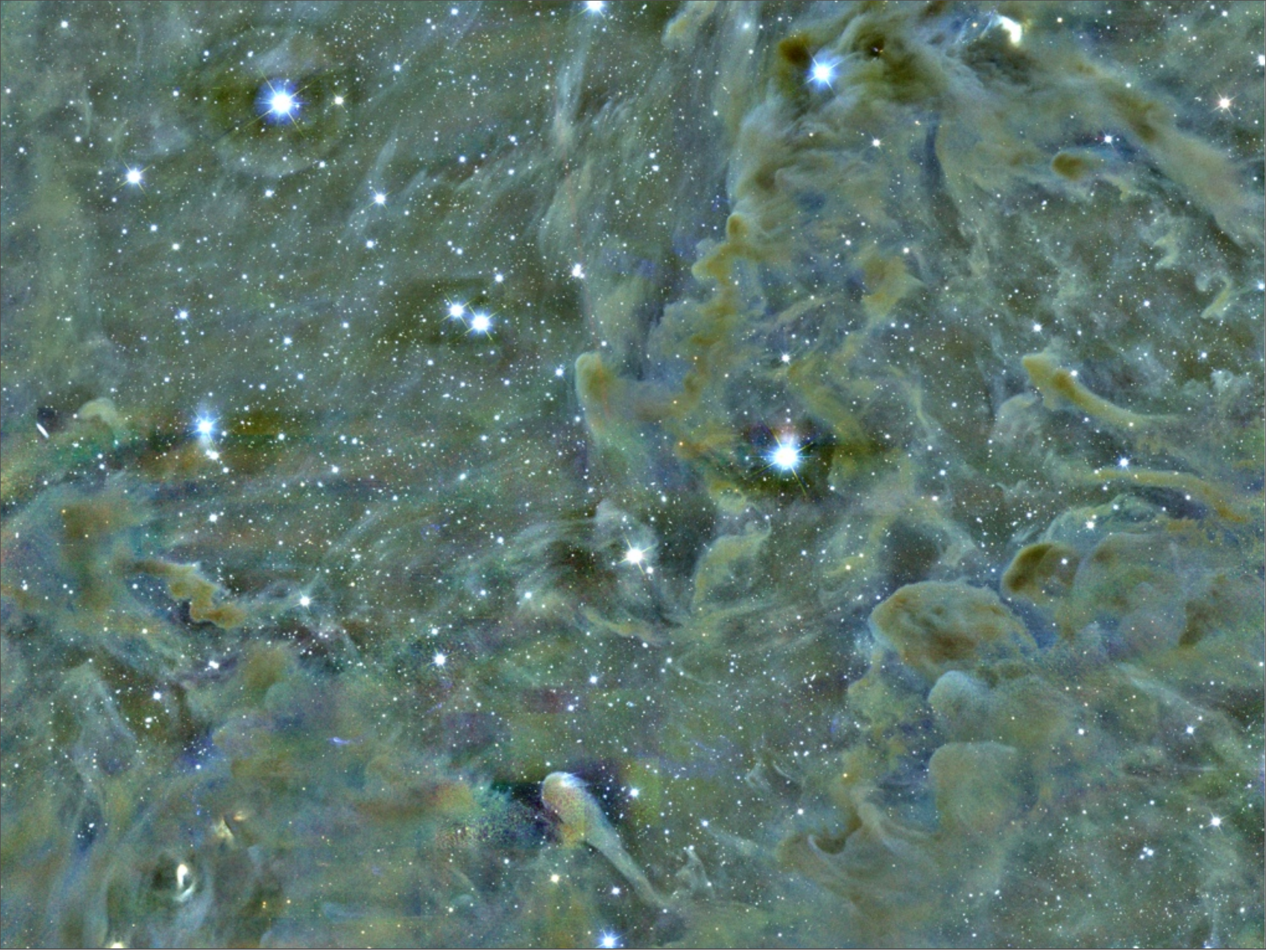


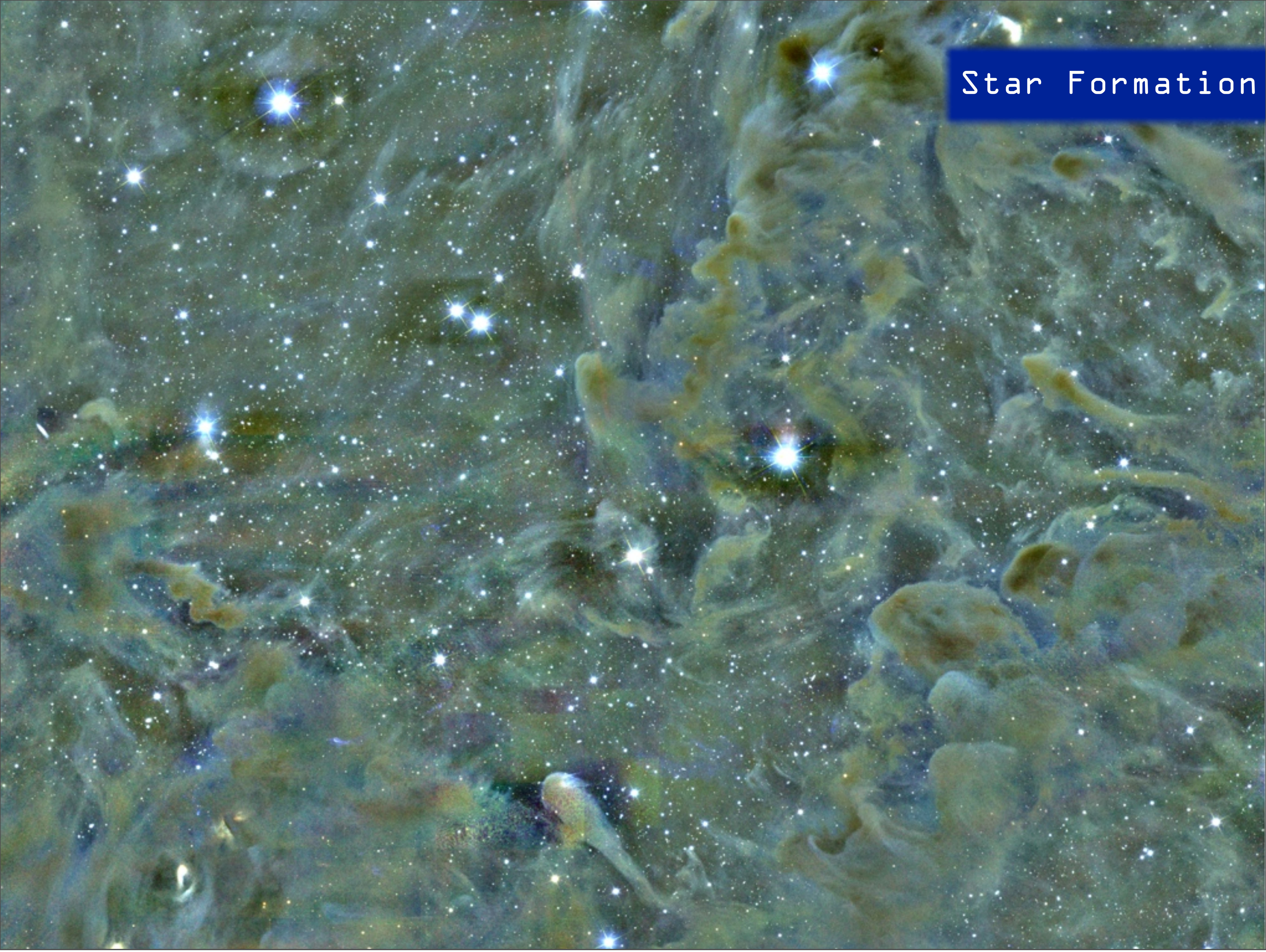
Star (and Planet, and Moon) Formation 101



Star (and Planet, and Moon) Formation 101






Star Formation

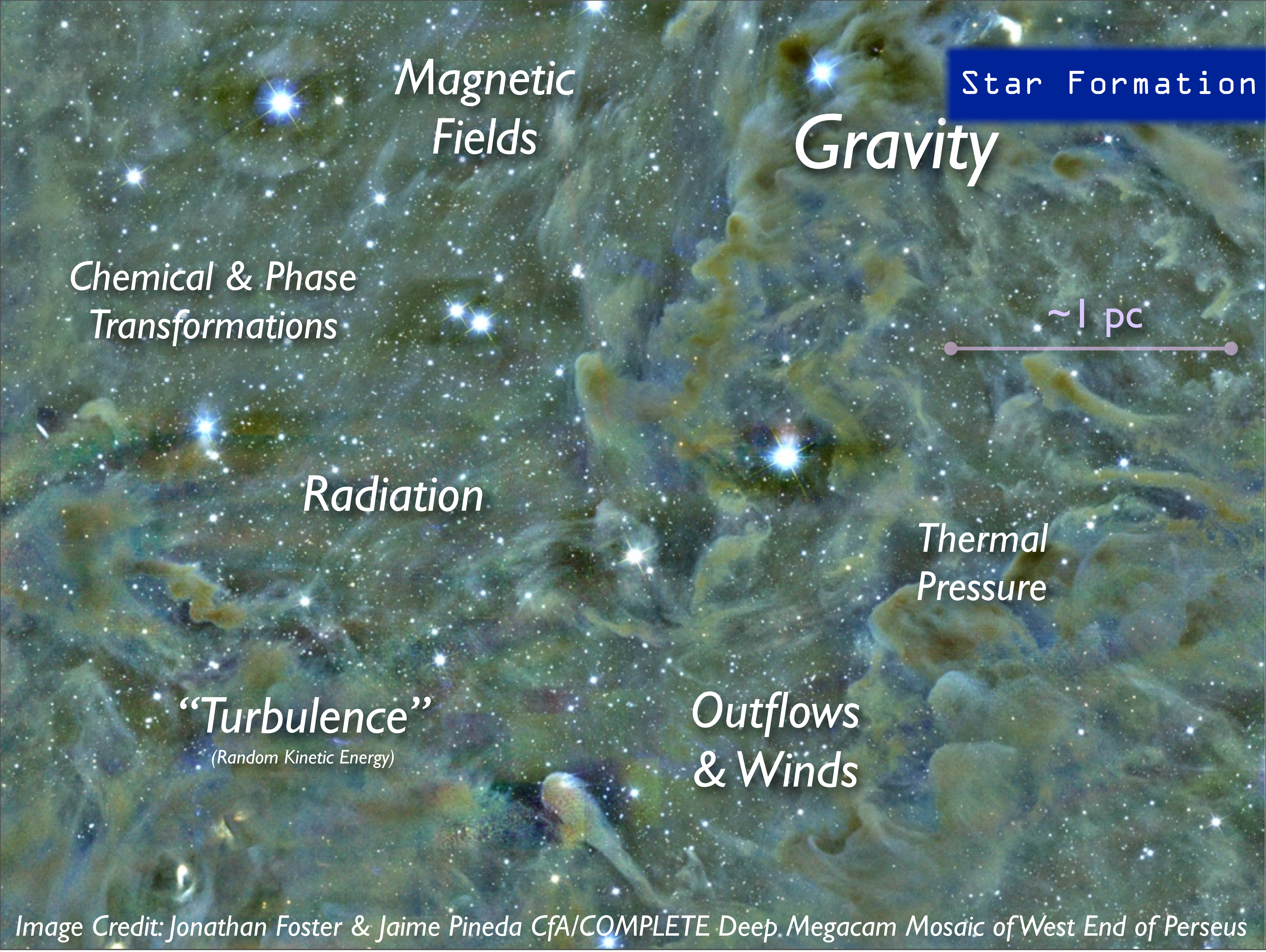
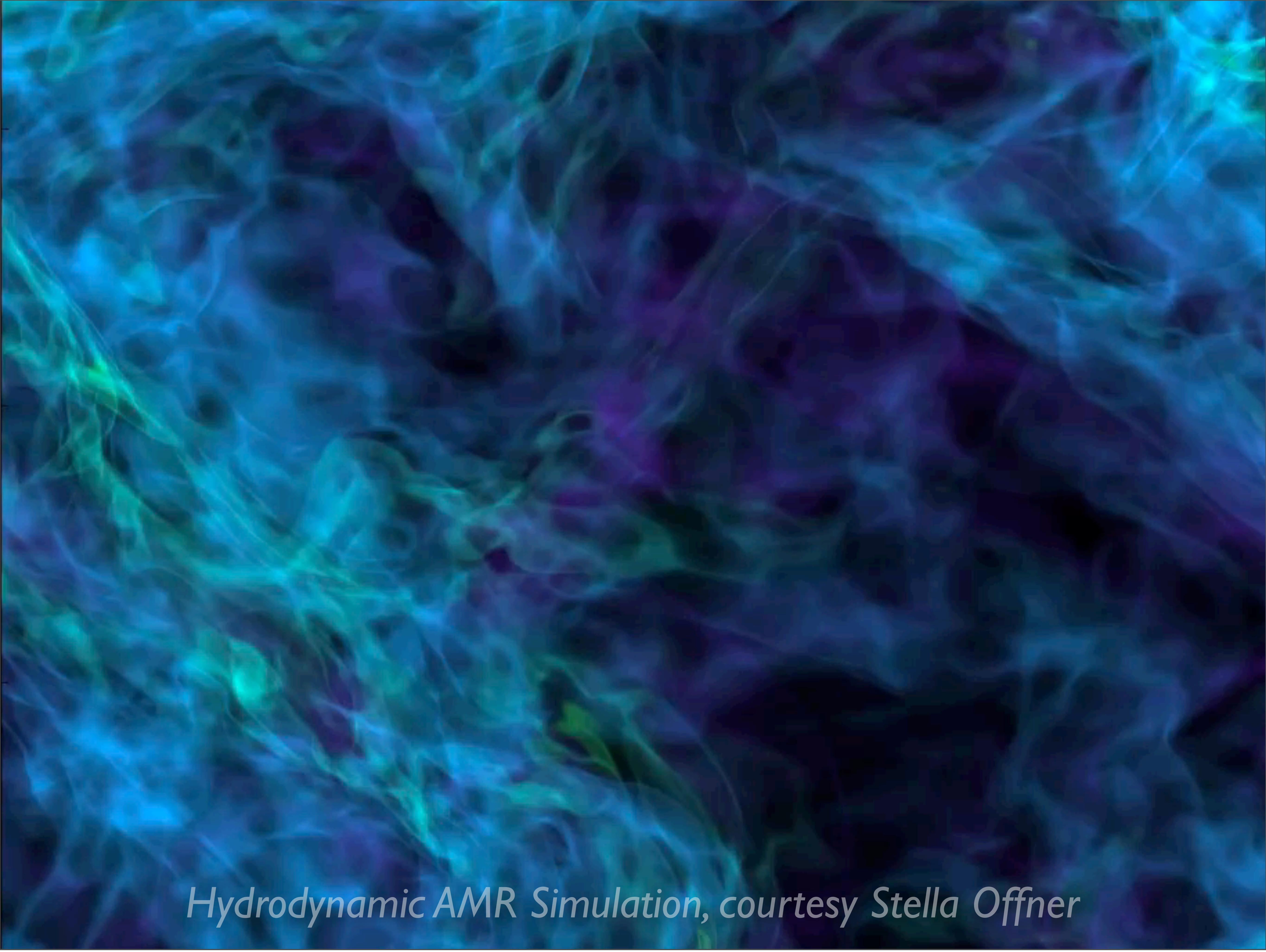
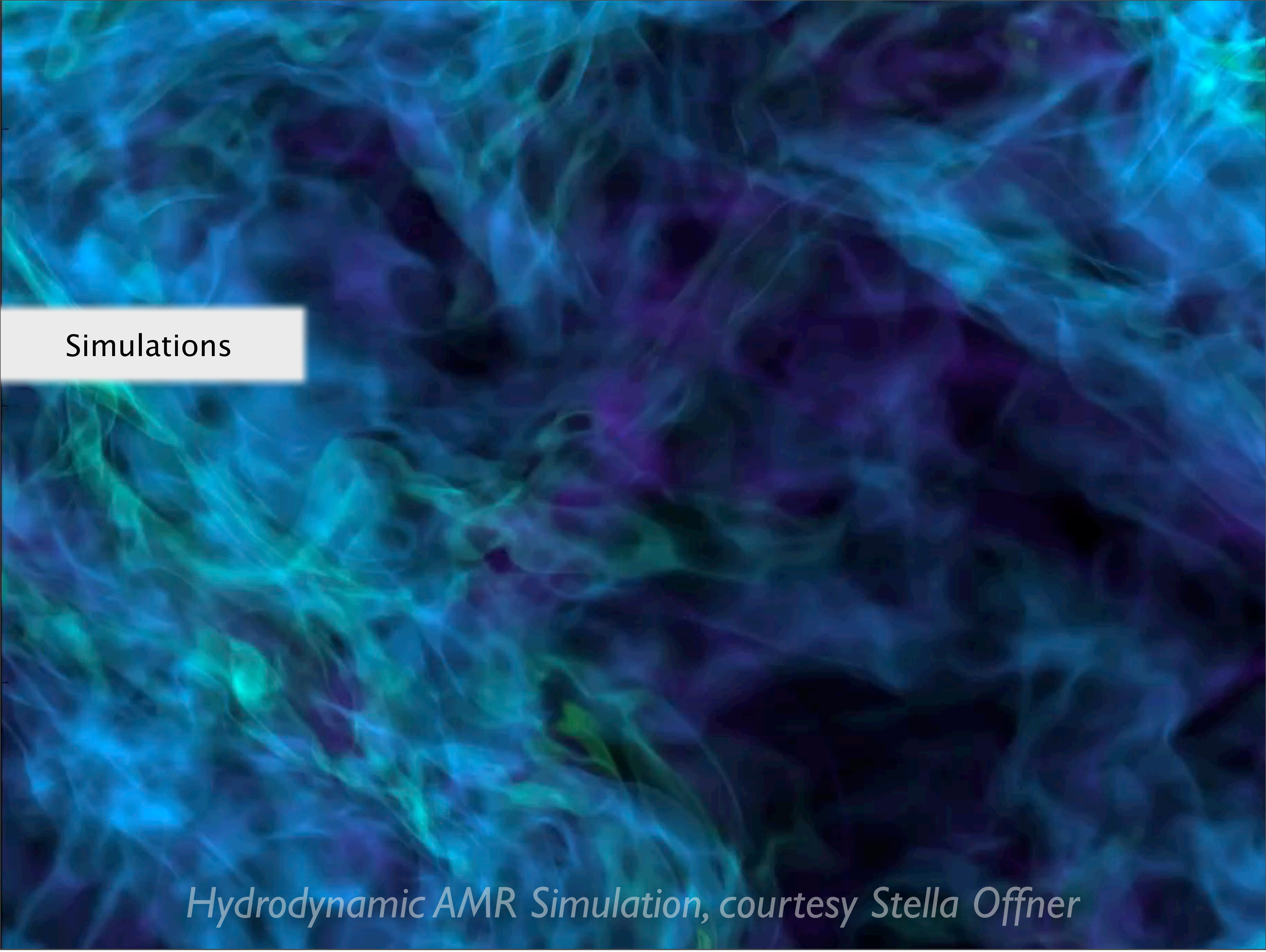


Image Credit: Jonathan Foster & Jaime Pineda CfA/COMPLETE Deep Megacam Mosaic of West End of Perseus



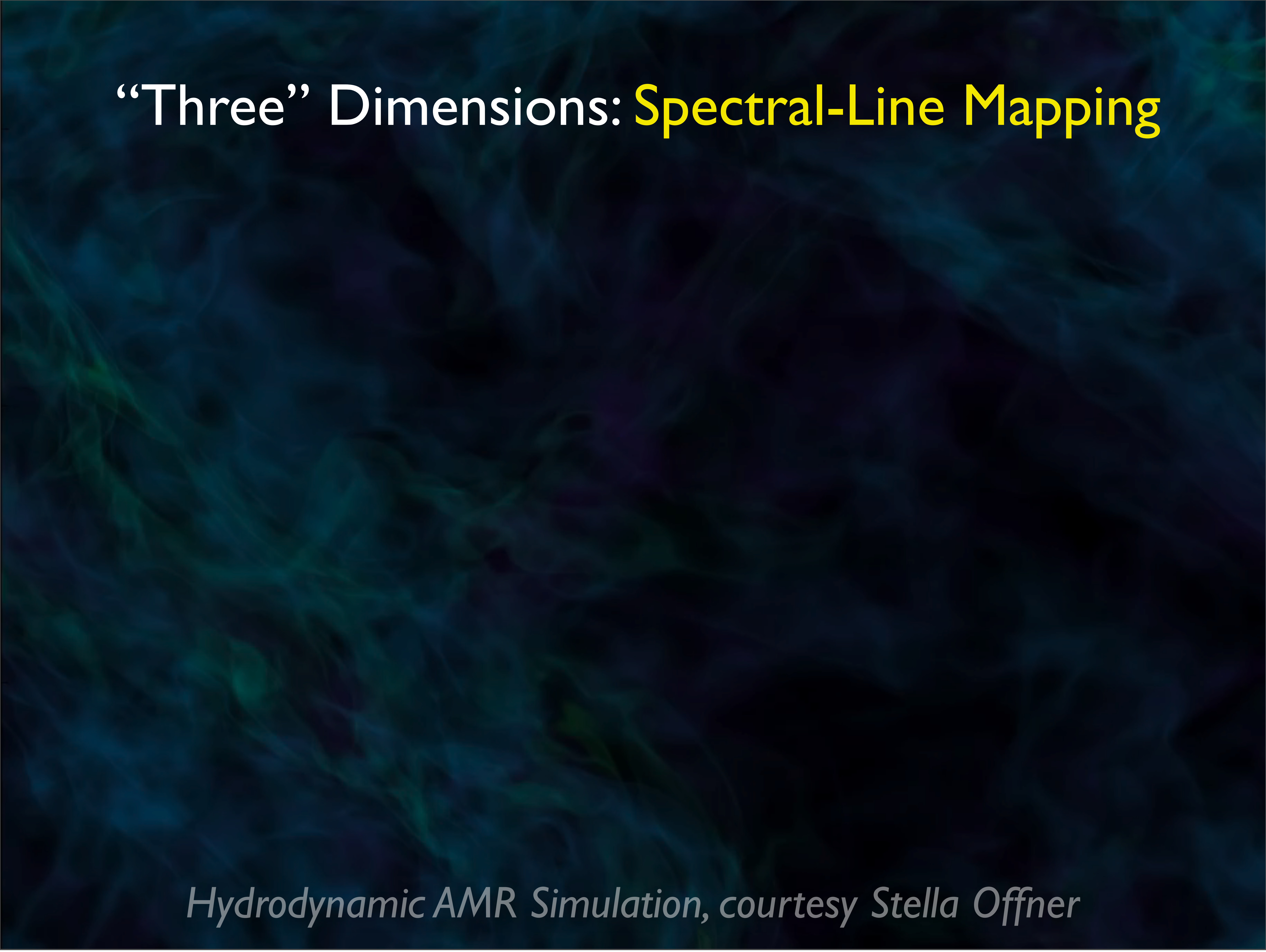
Hydrodynamic AMR Simulation, courtesy Stella Offner

Simulations



Hydrodynamic AMR Simulation, courtesy Stella Offner

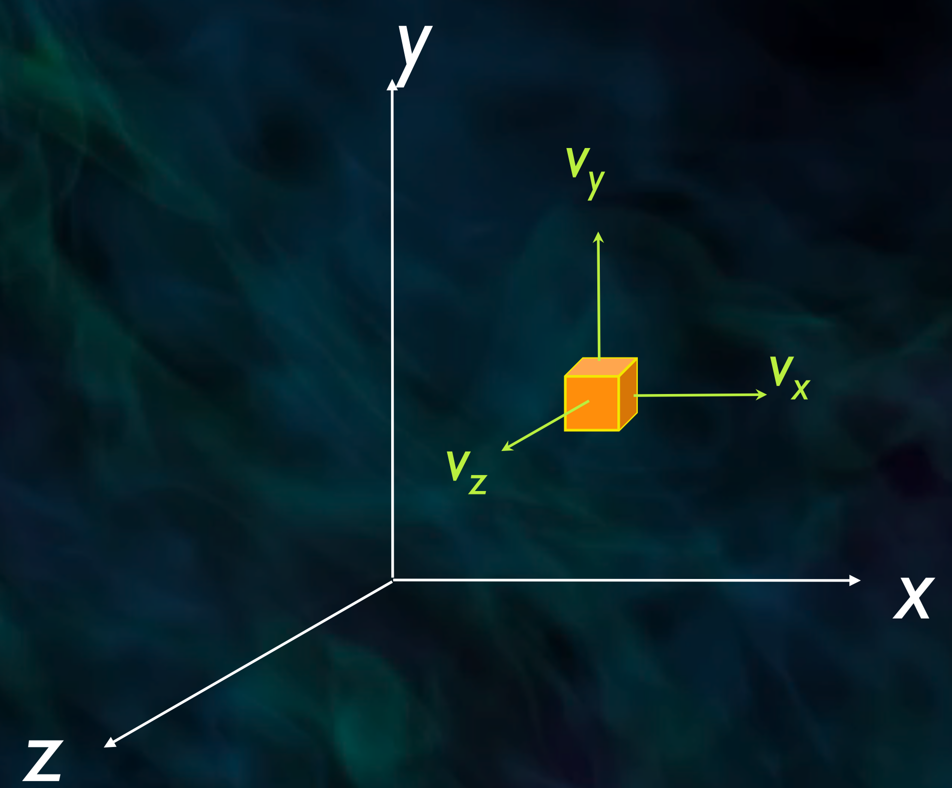
“Three” Dimensions: Spectral-Line Mapping



Hydrodynamic AMR Simulation, courtesy Stella Offner

“Three” Dimensions: Spectral-Line Mapping

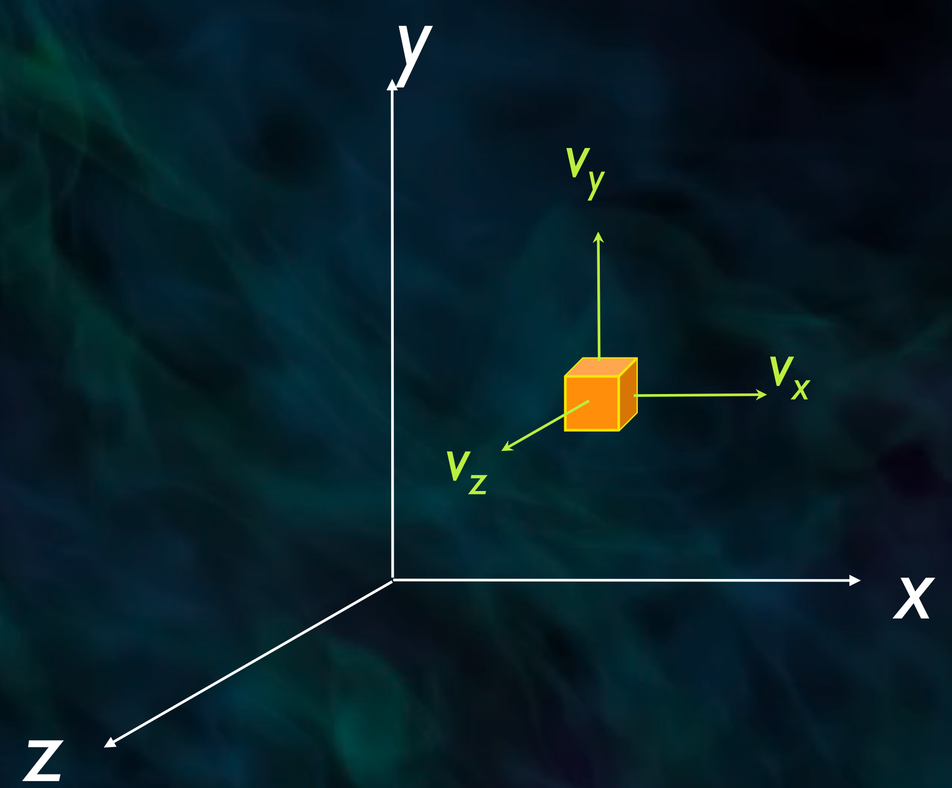
We wish we could measure…



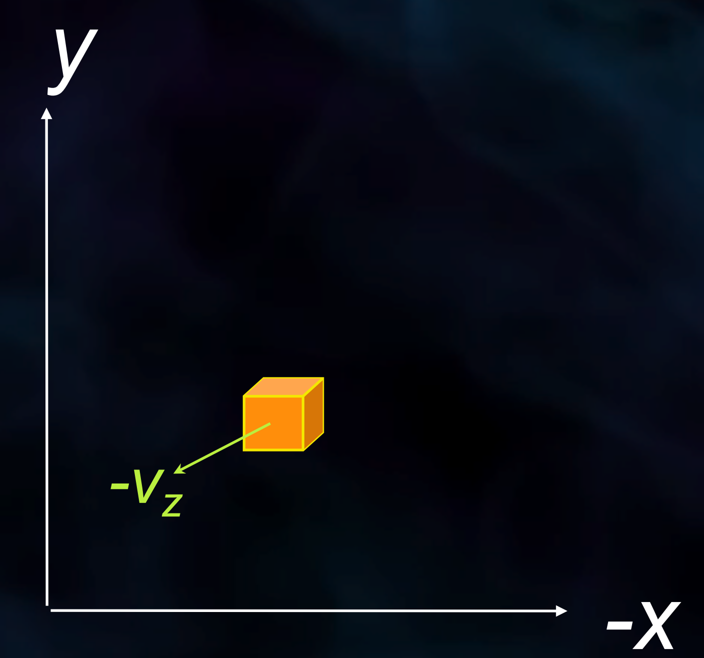
Hydrodynamic AMR Simulation, courtesy Stella Offner

"Three" Dimensions: Spectral-Line Mapping

We wish we could measure…



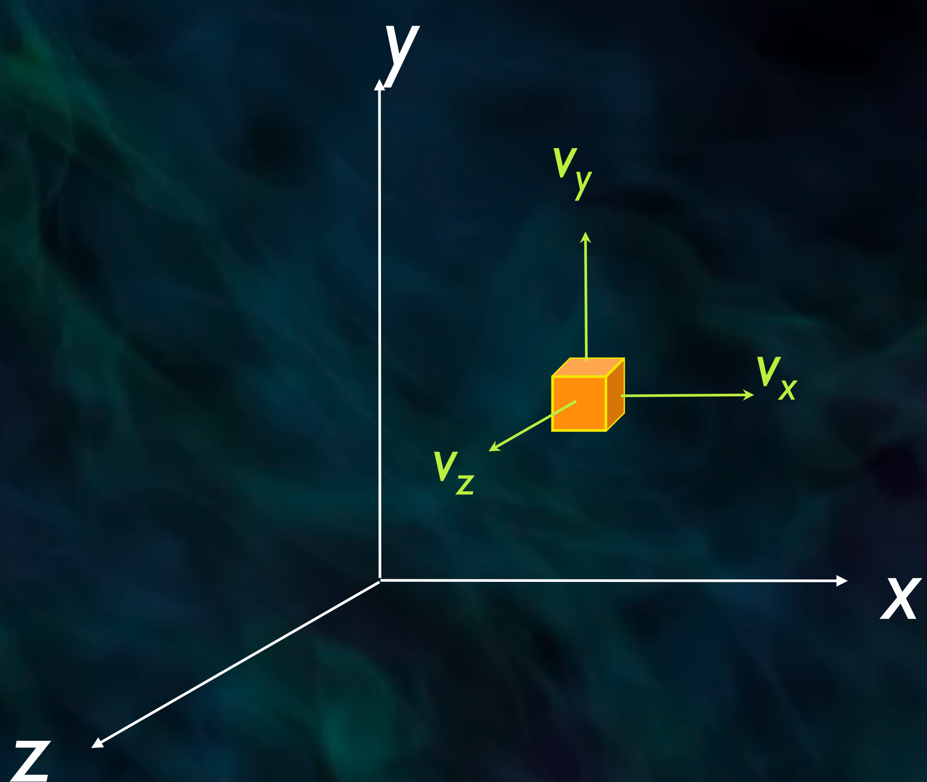
But we can measure…



Hydrodynamic AMR Simulation, courtesy Stella Offner

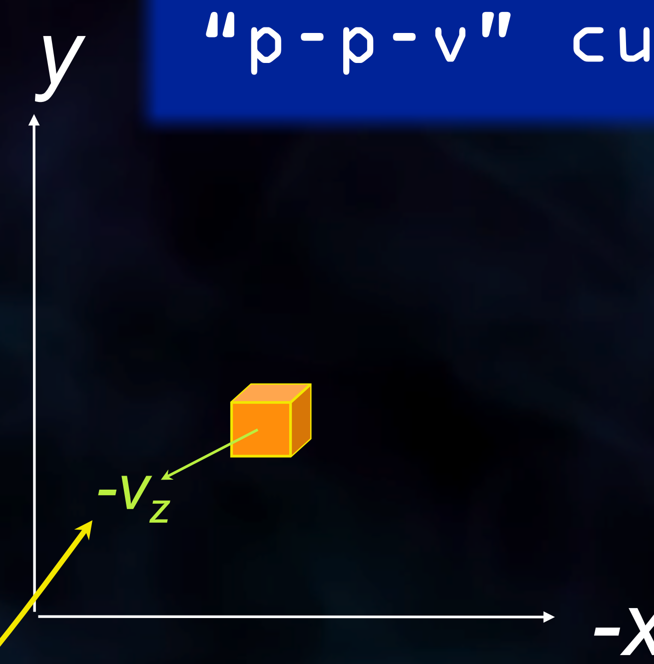
"Three" Dimensions: Spectral-Line Mapping

We wish we could measure…



But we can measure…

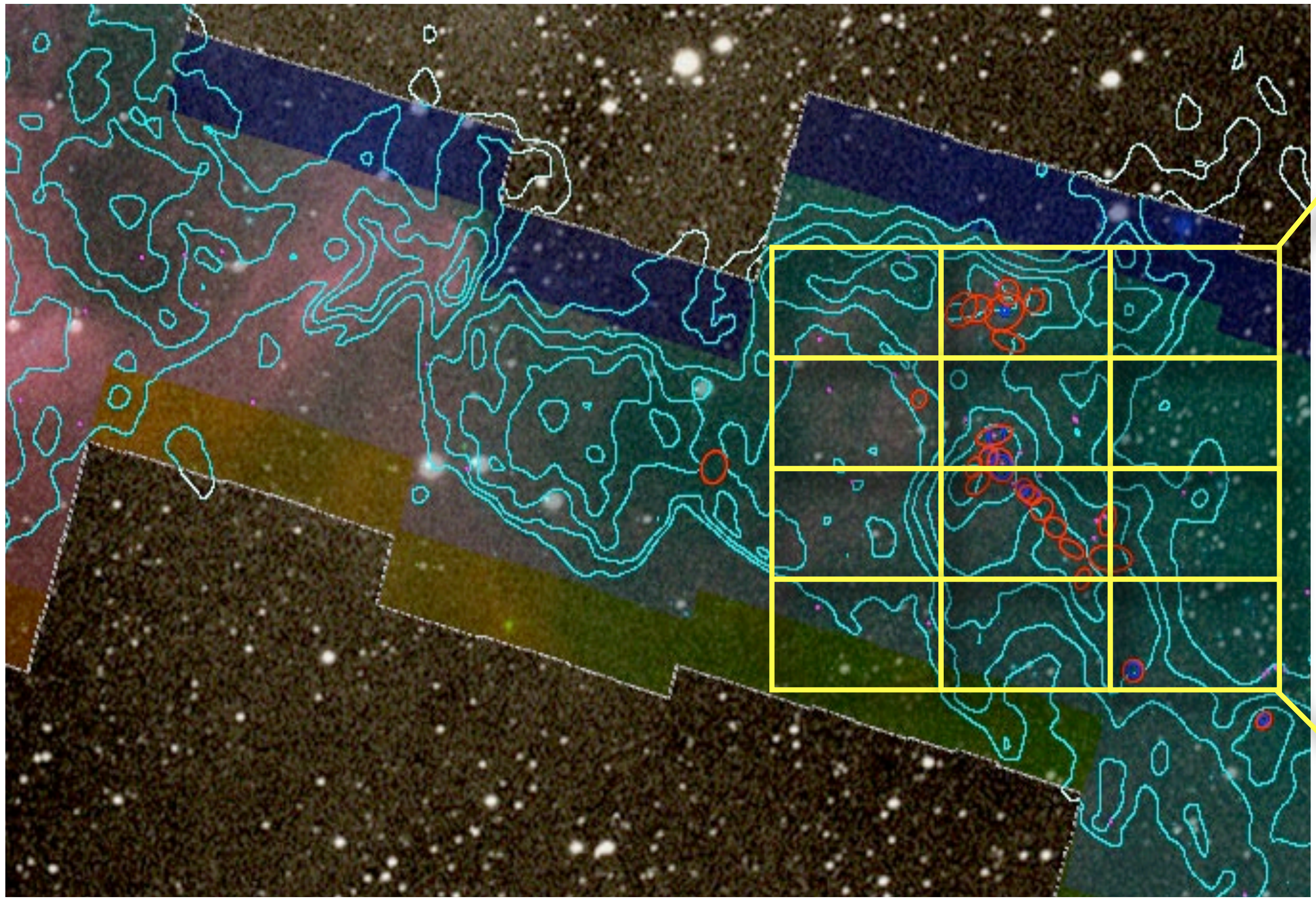
"p-p-v" cubes

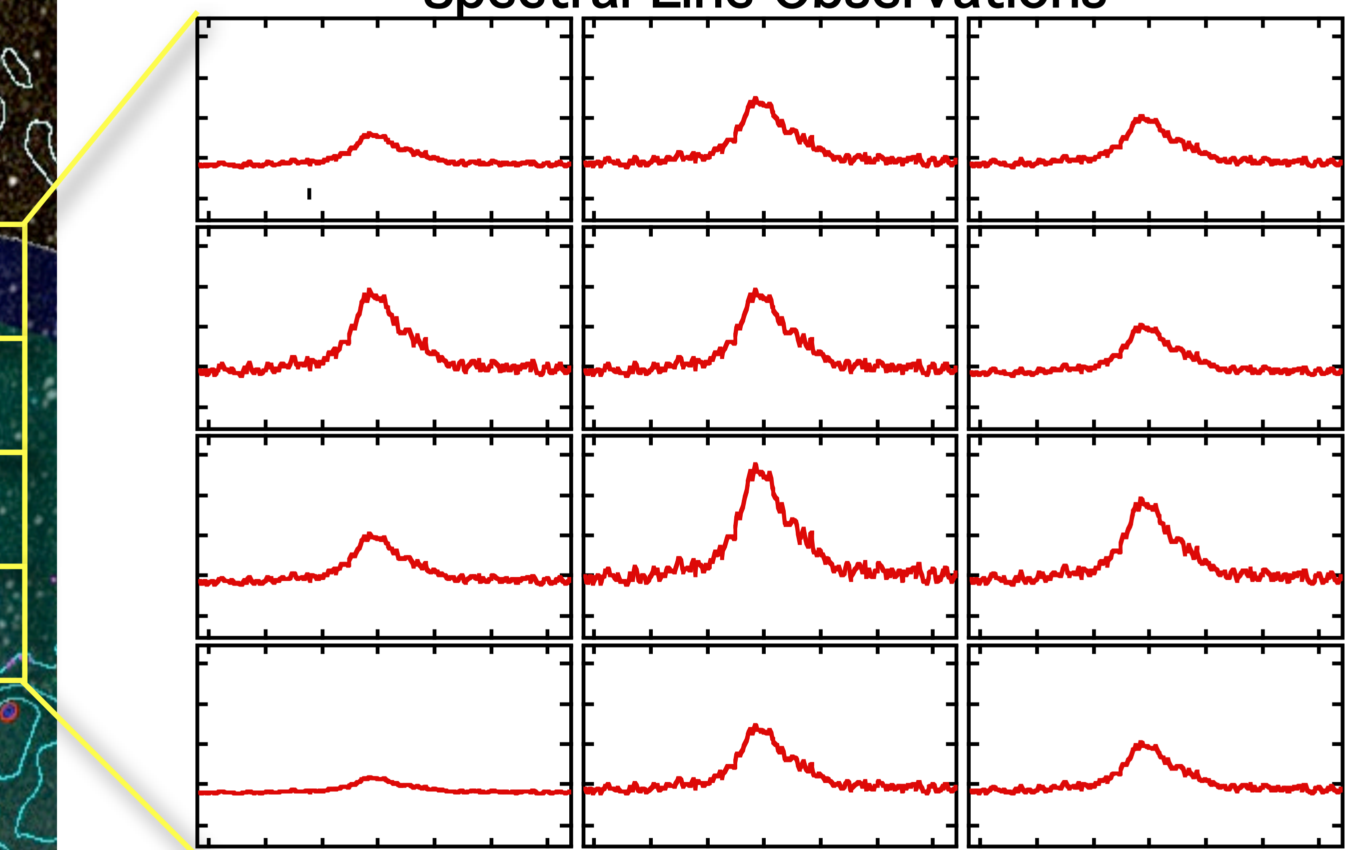


v_z *only* from "spectral-line maps"

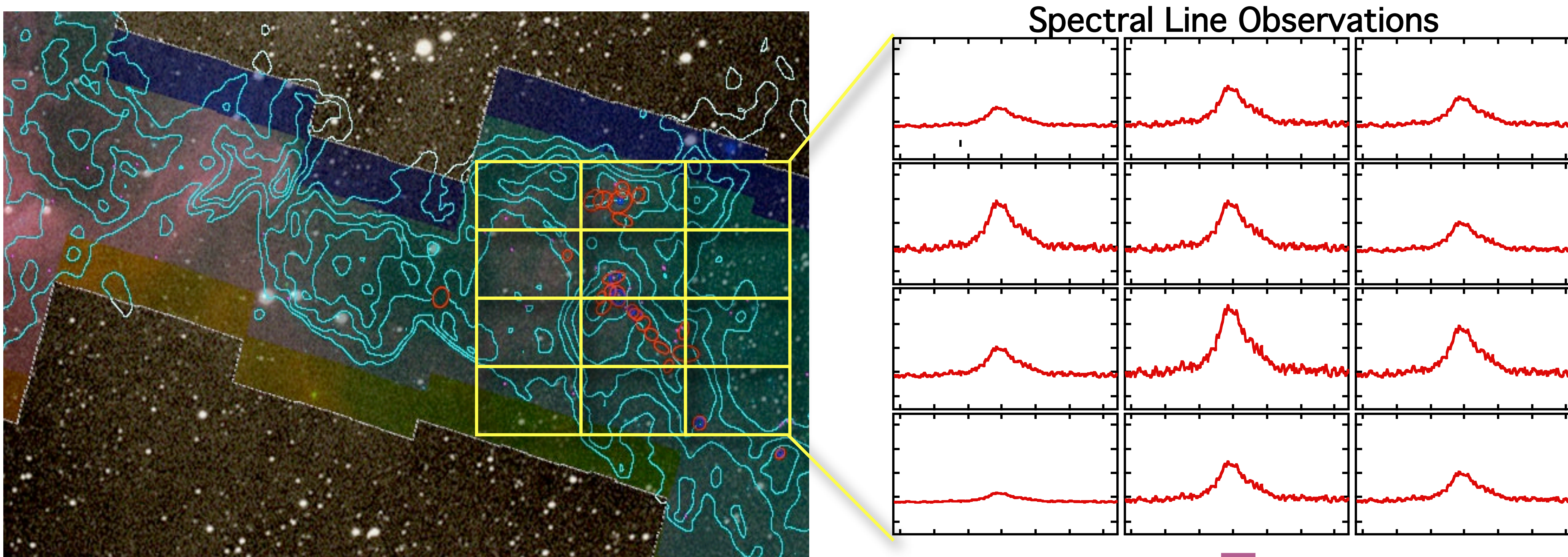
Hydrodynamic AMR Simulation, courtesy Stella Offner

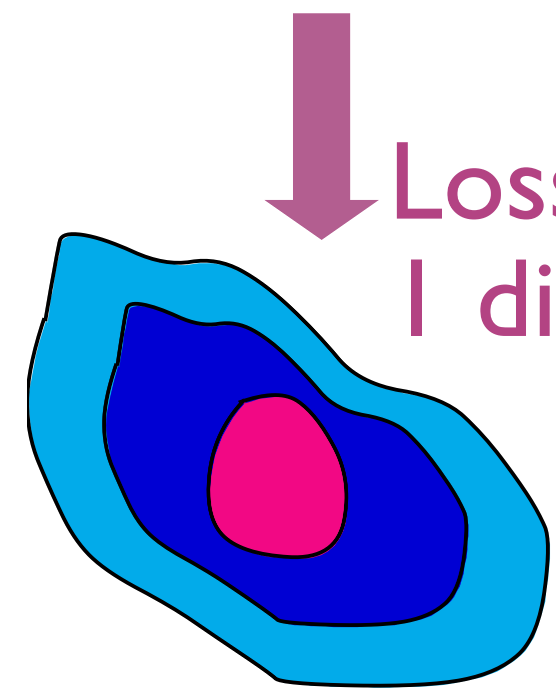
Spectral-Line Mapping





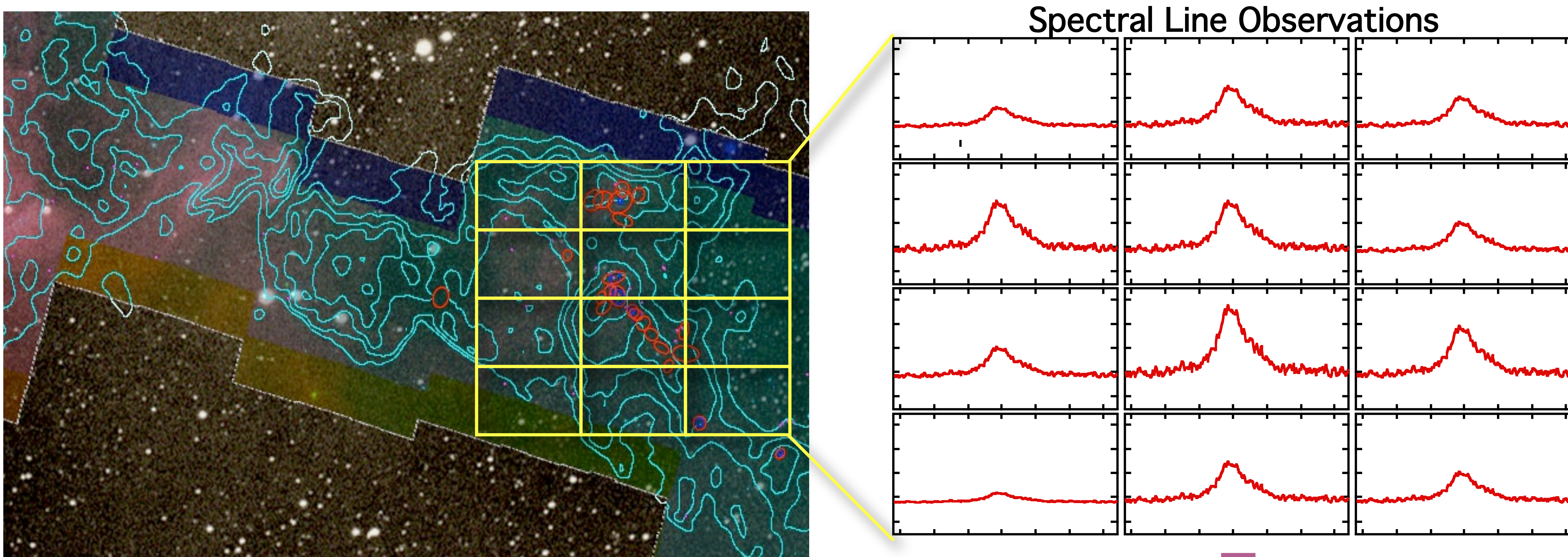
Spectral-Line Mapping

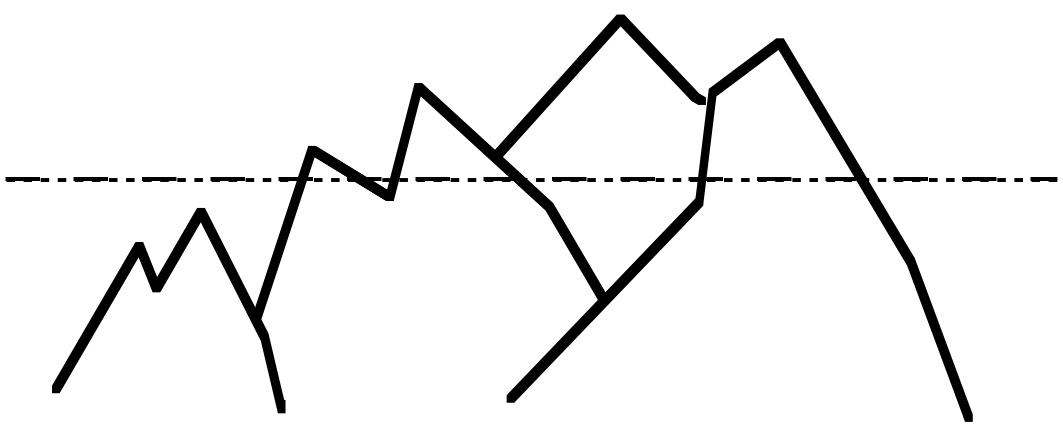




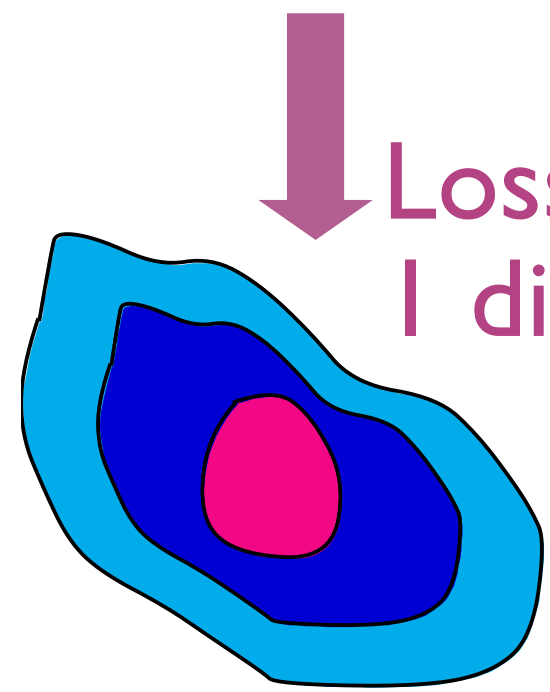
Loss of
1 dimension

Spectral-Line Mapping



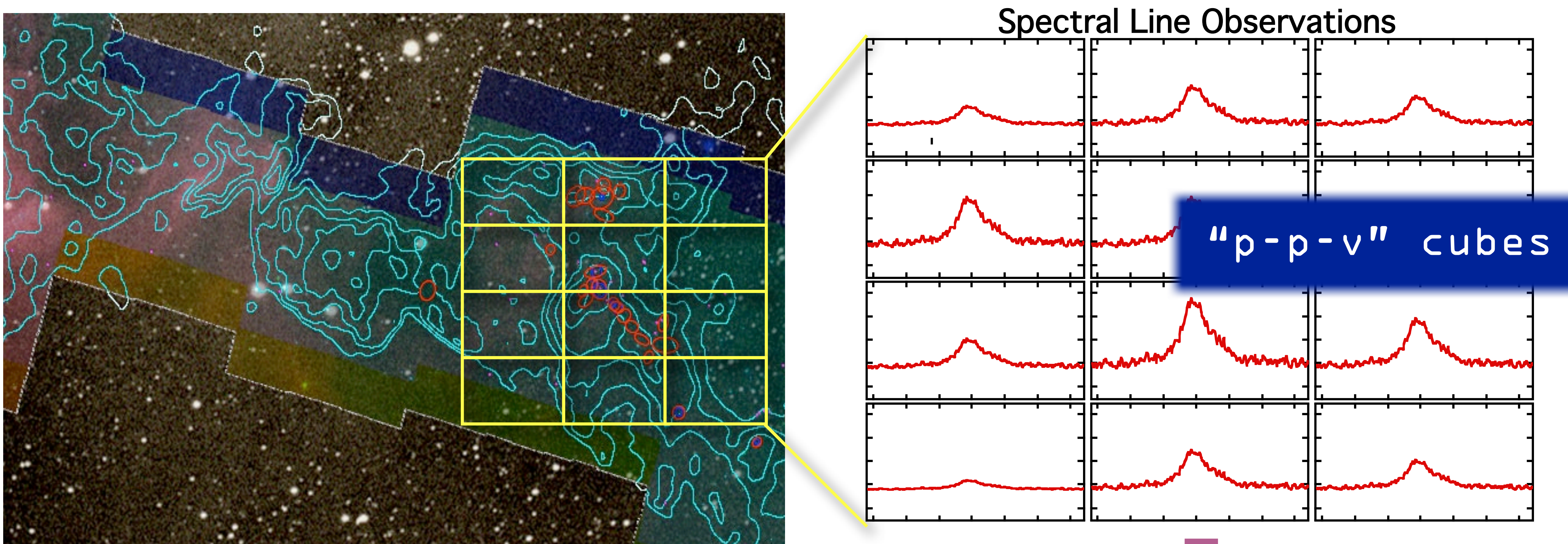


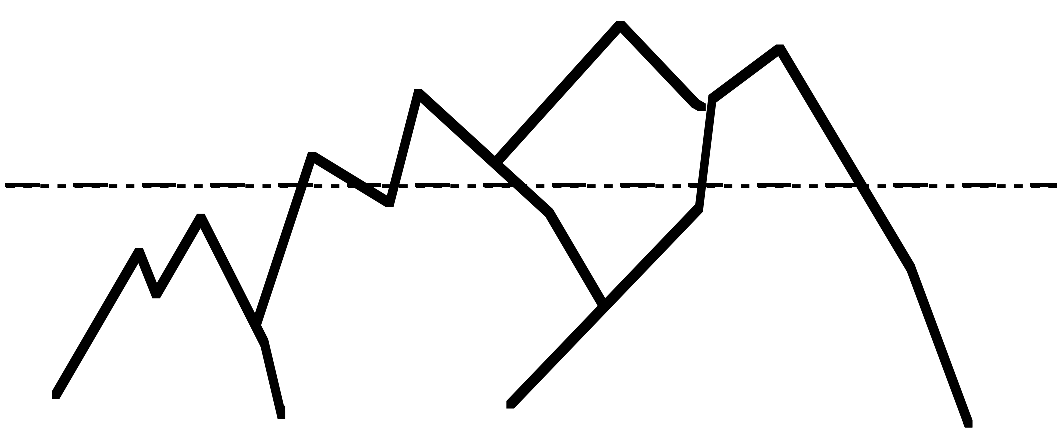
No loss of information



Loss of 1 dimension

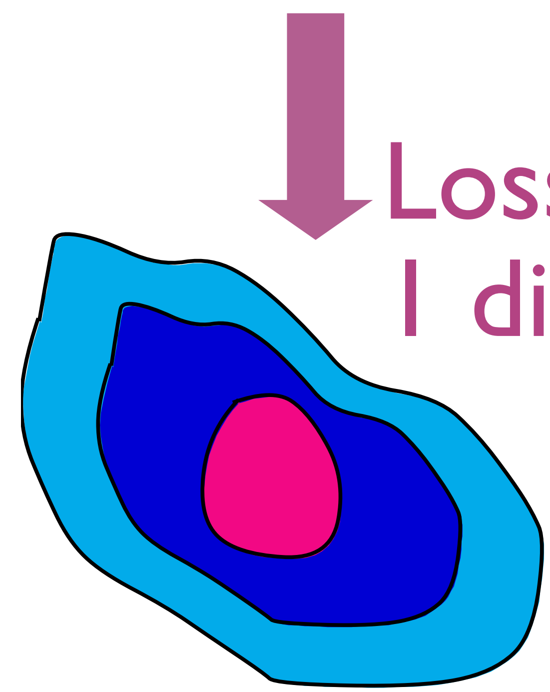
Spectral-Line Mapping





Mountain Range

No loss of information

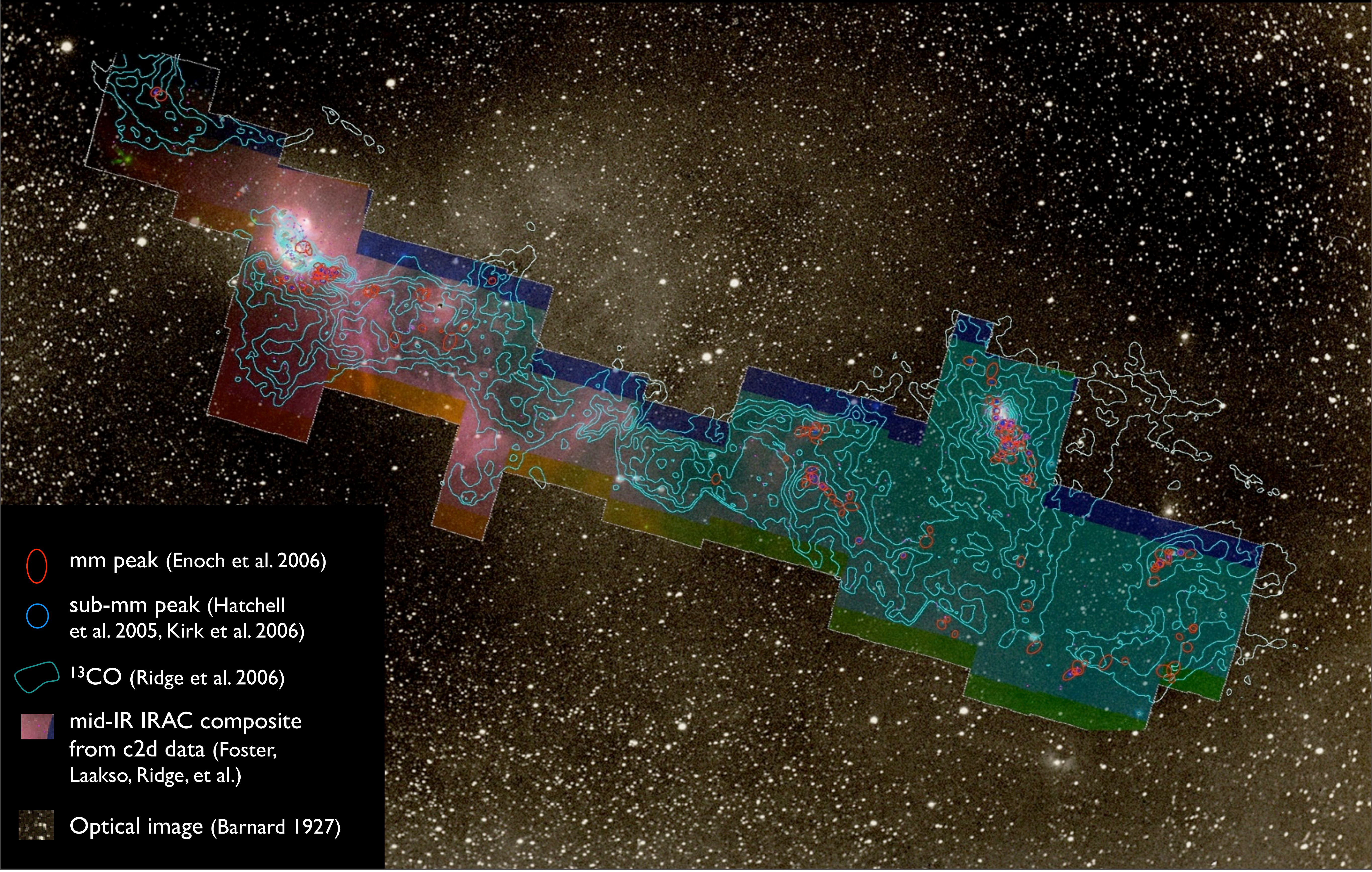


Loss of 1 dimension

High-Dimensional Data

COMPLETE Perseus

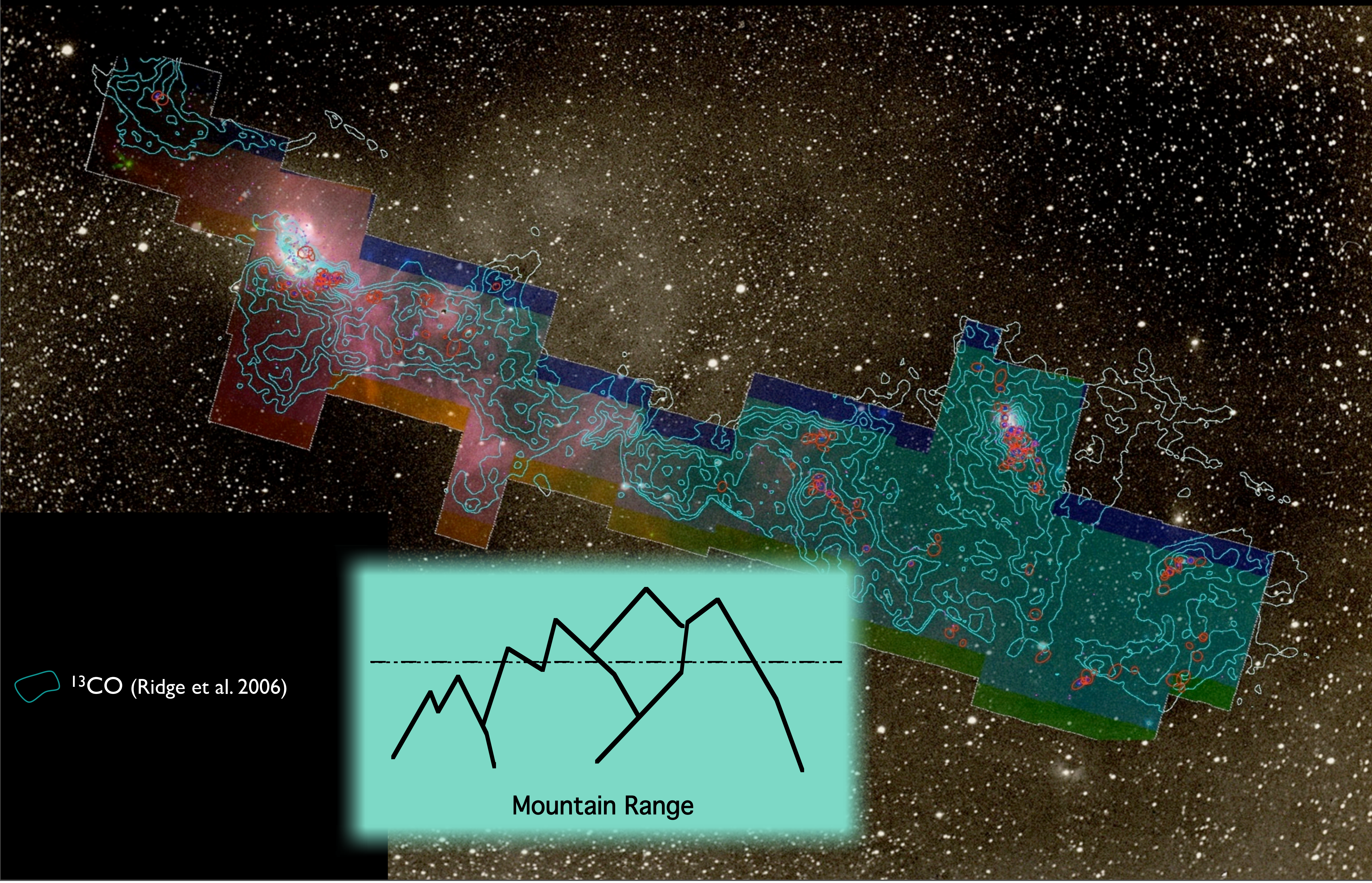
Star Formation



High-Dimensional Data

COMPLETE Perseus

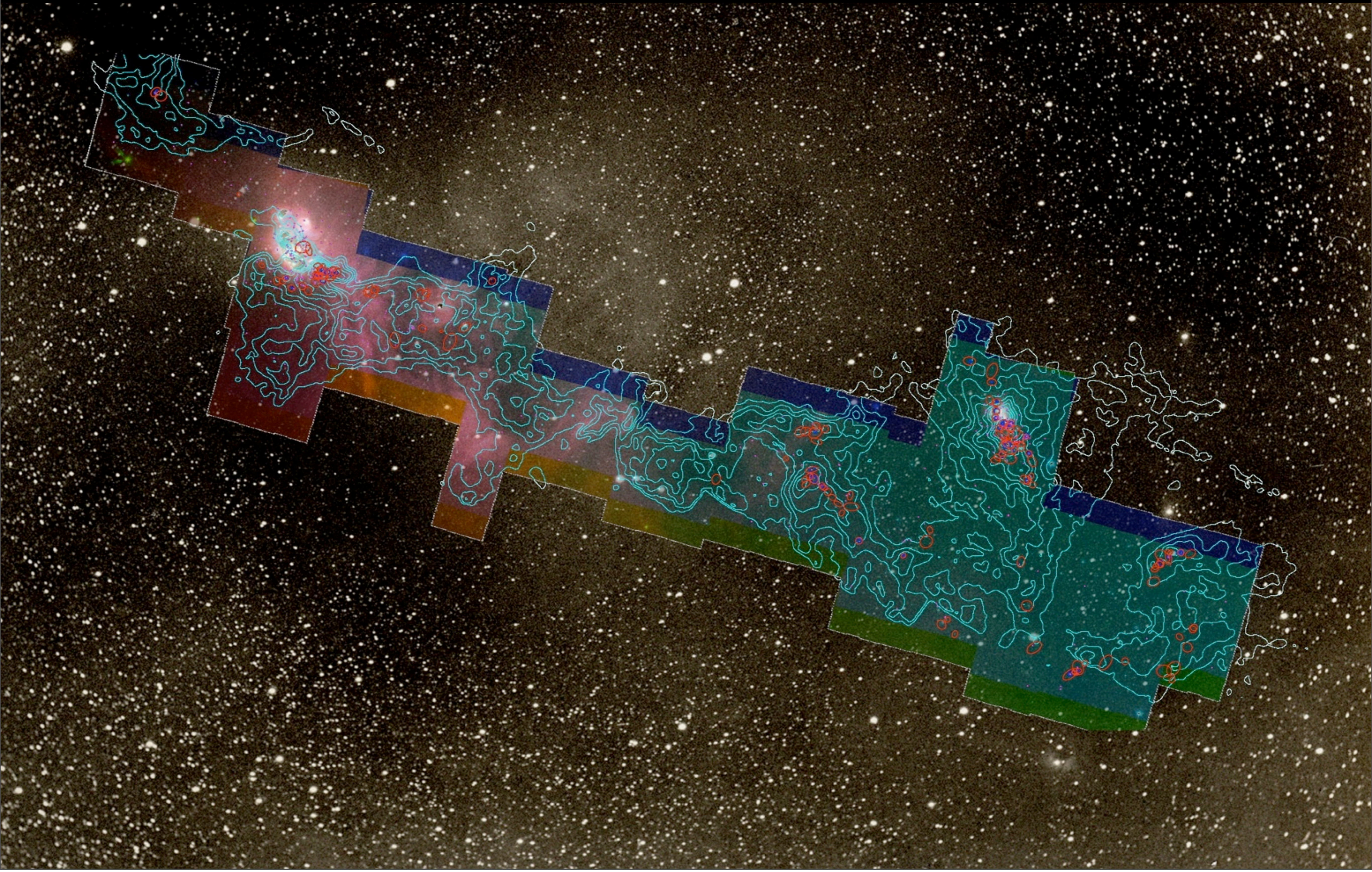
Star Formation



High-Dimensional Data

COMPLETE Perseus

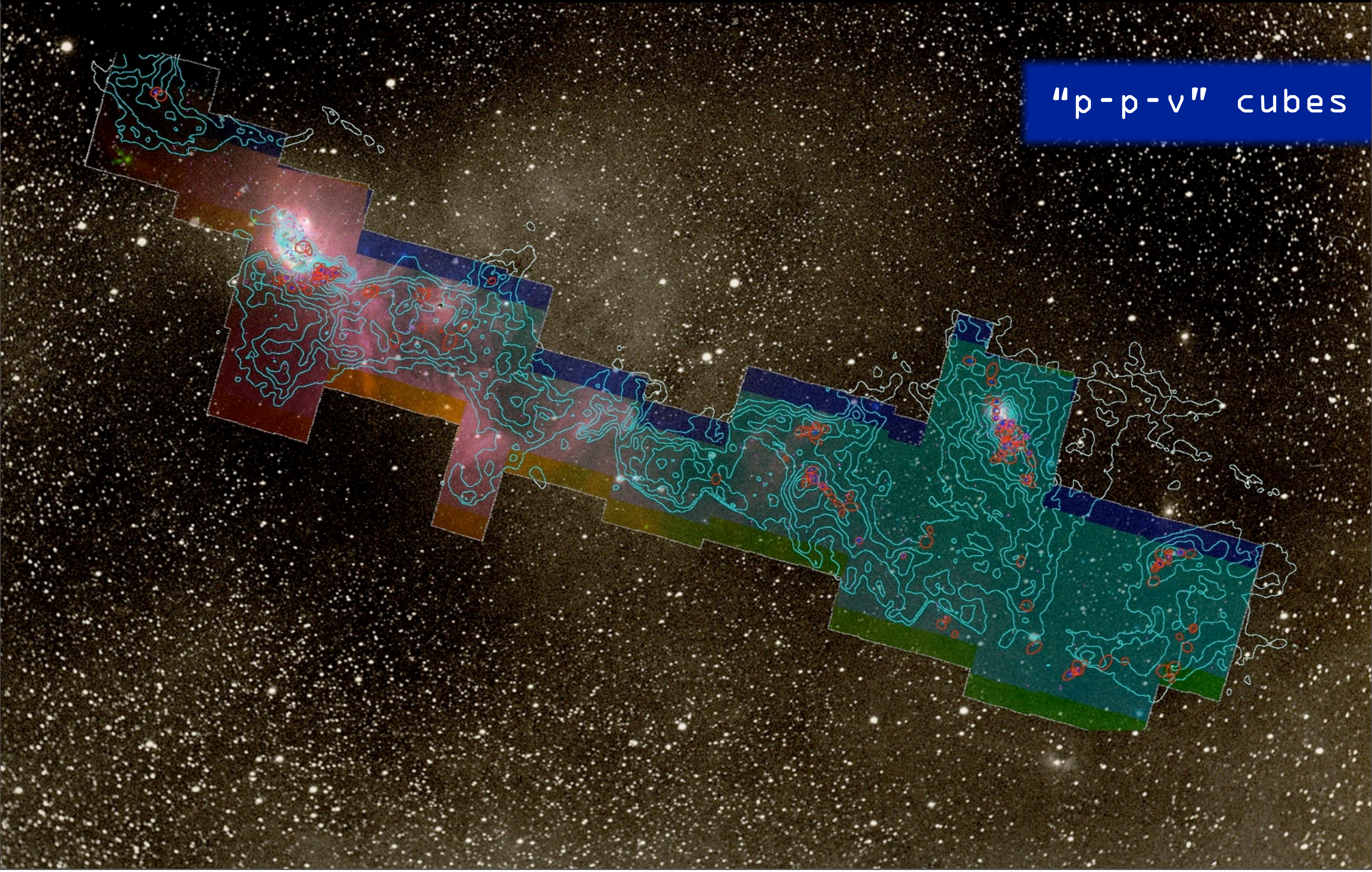
Star Formation



High-Dimensional Data

COMPLETE Perseus

Star Formation

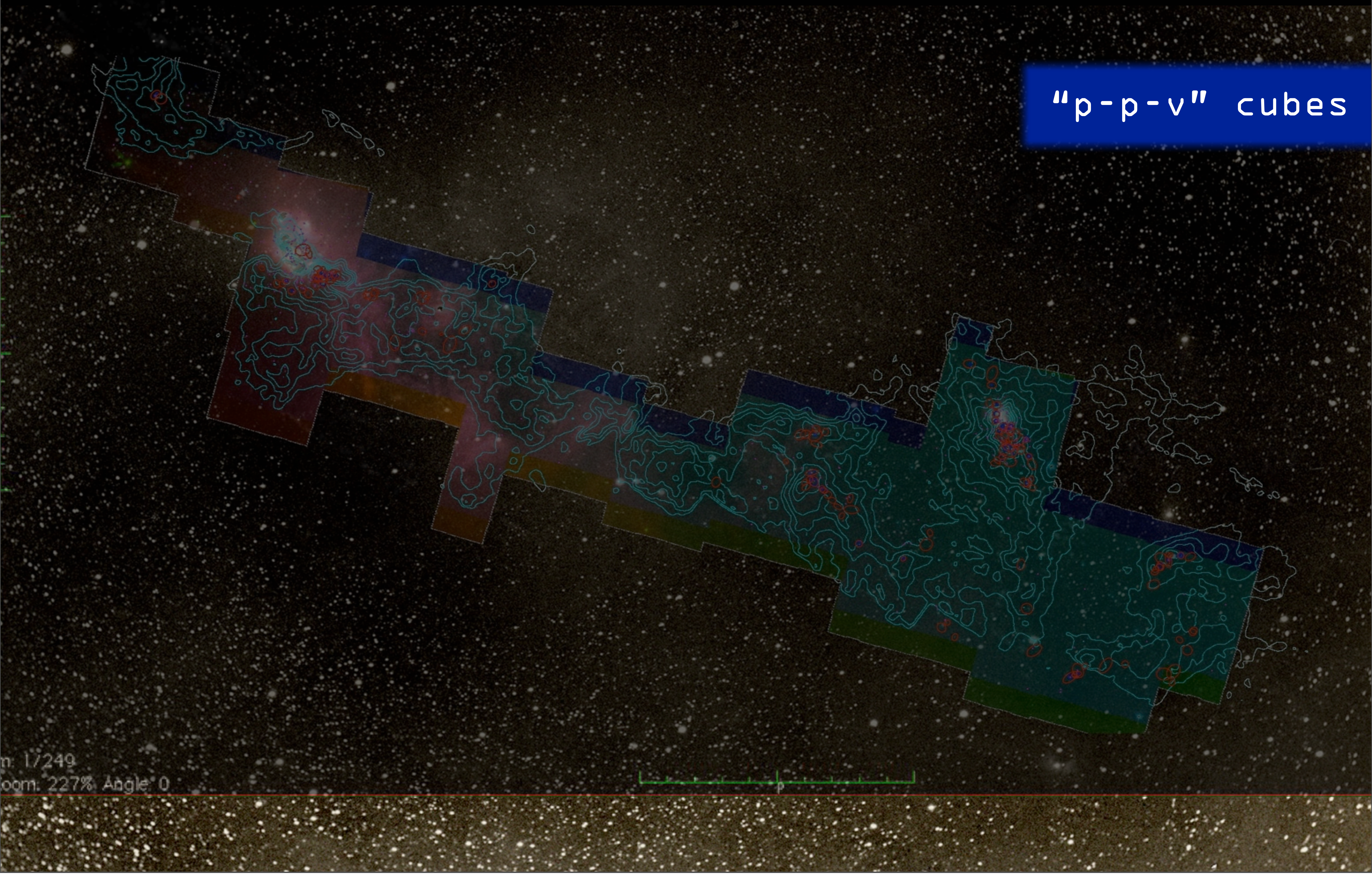


High-Dimensional Data

COMPLETE Perseus

Star Formation

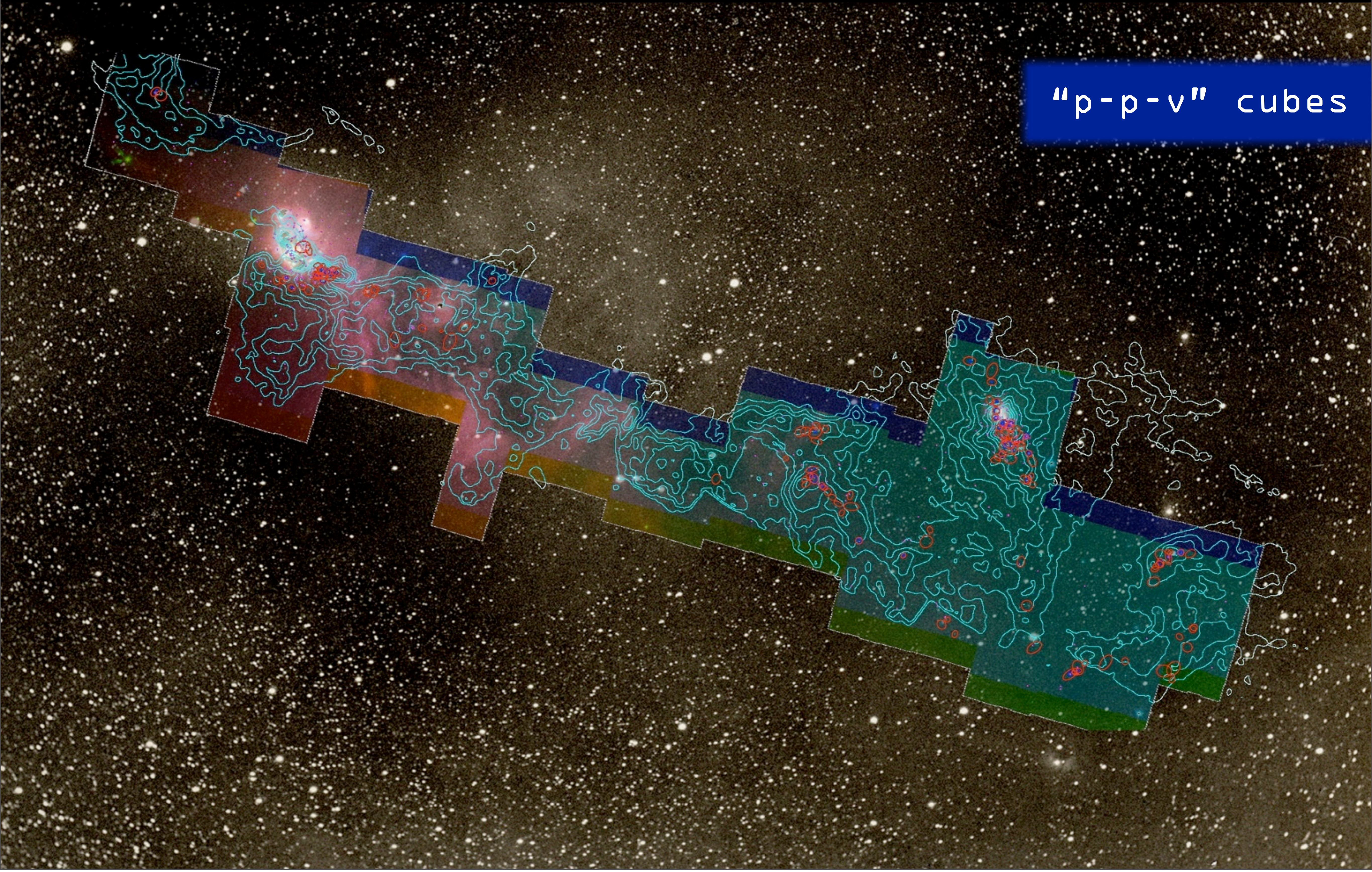
"p-p-v" cubes



High-Dimensional Data

COMPLETE Perseus

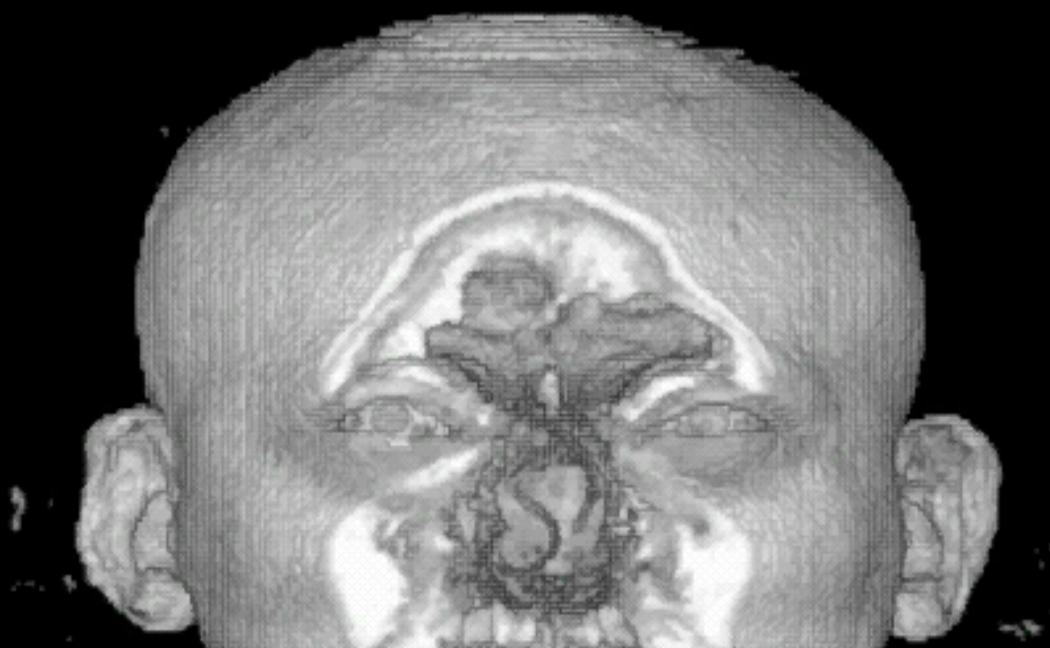
Star Formation



"p-p-v" cubes

“Astronomical Medicine”

“KEITH”



“z” is depth into head

“PERSEUS”

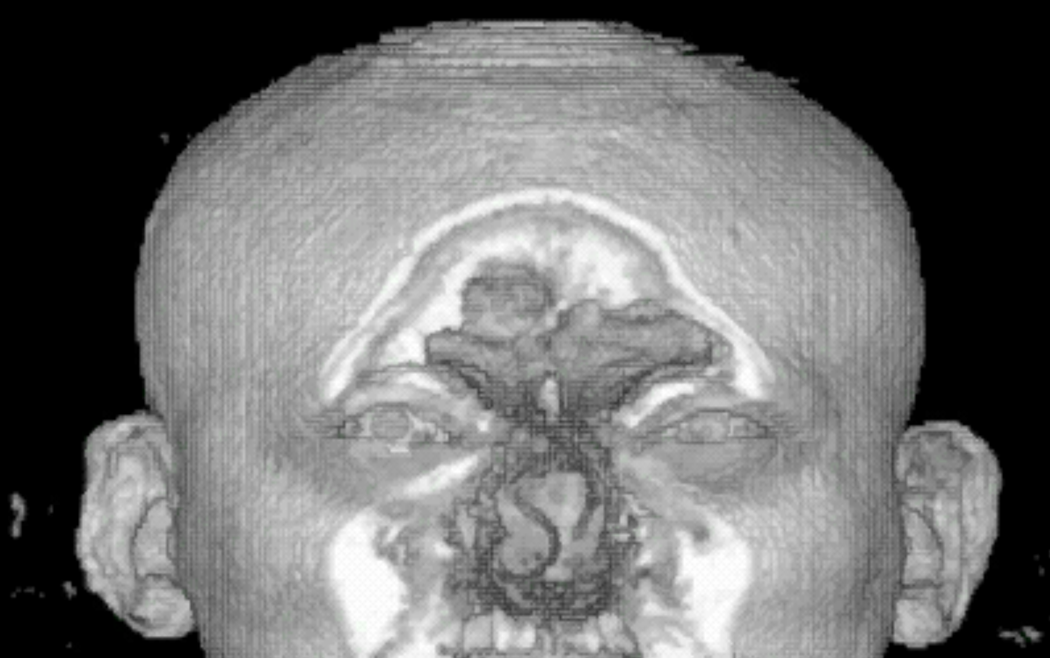
“z” is line-of-sight velocity

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

“Astronomical Medicine”

“KEITH”

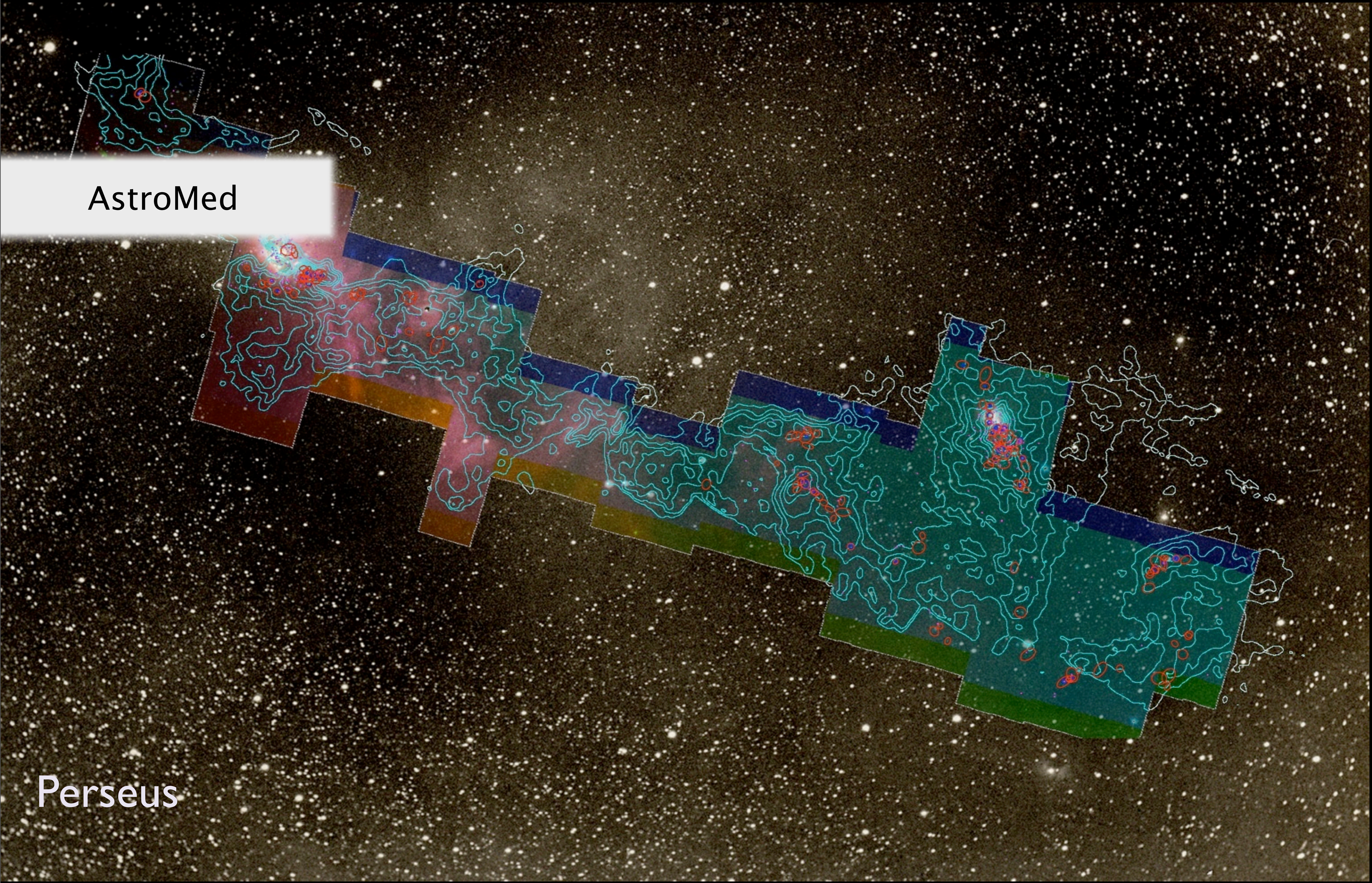
“PERSEUS”

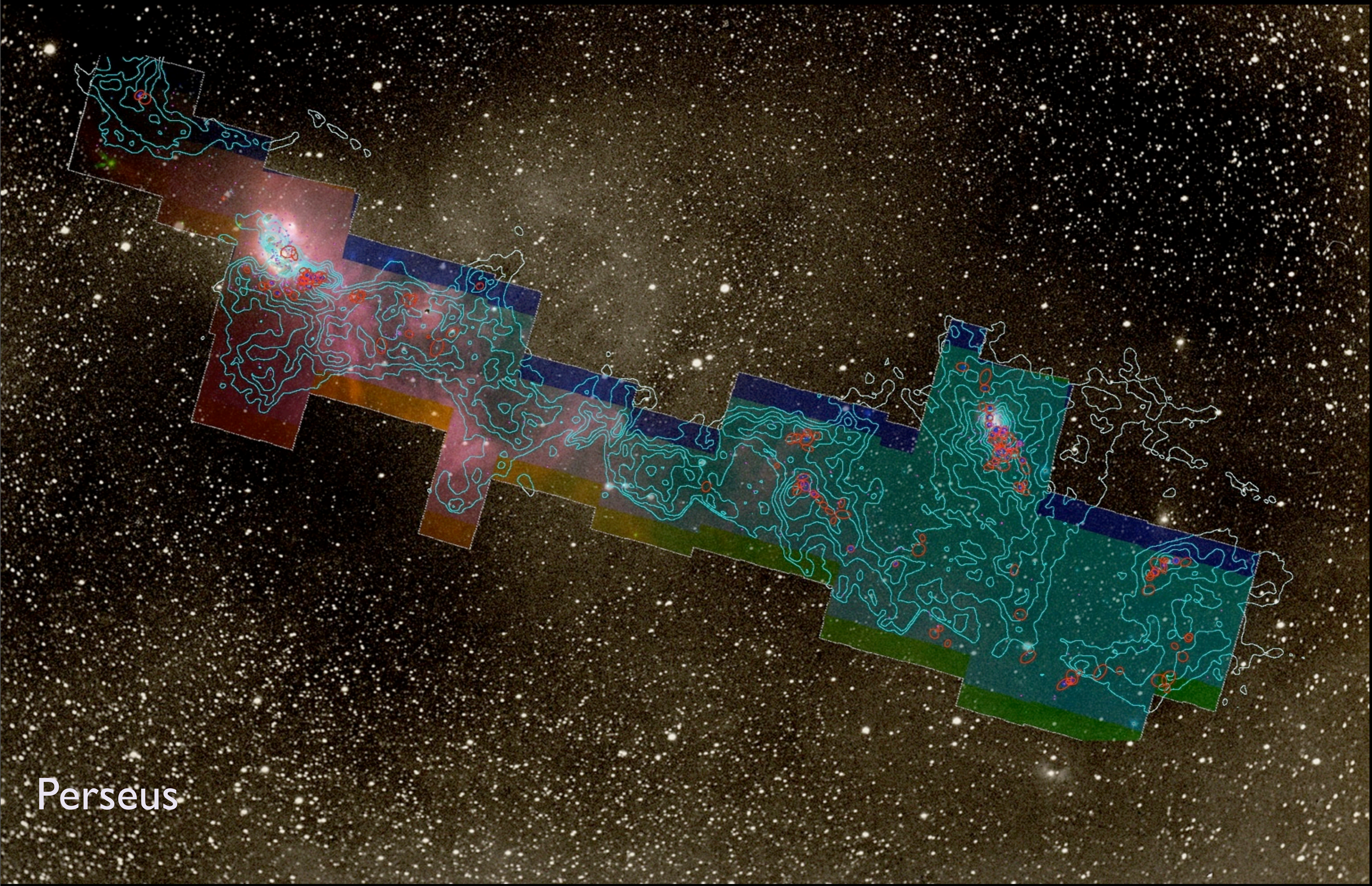


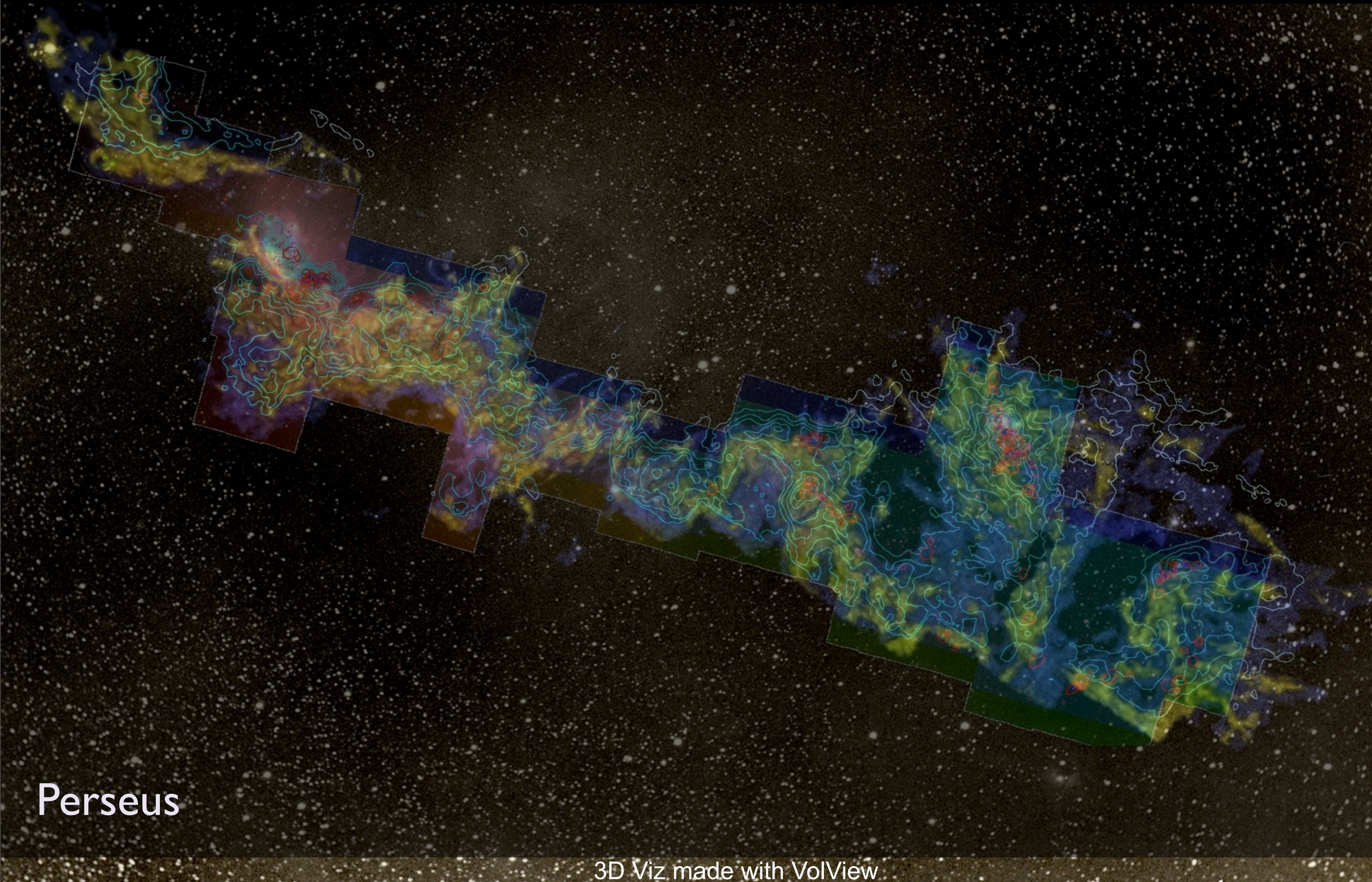
“z” is depth into head

“z” is line-of-sight velocity

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

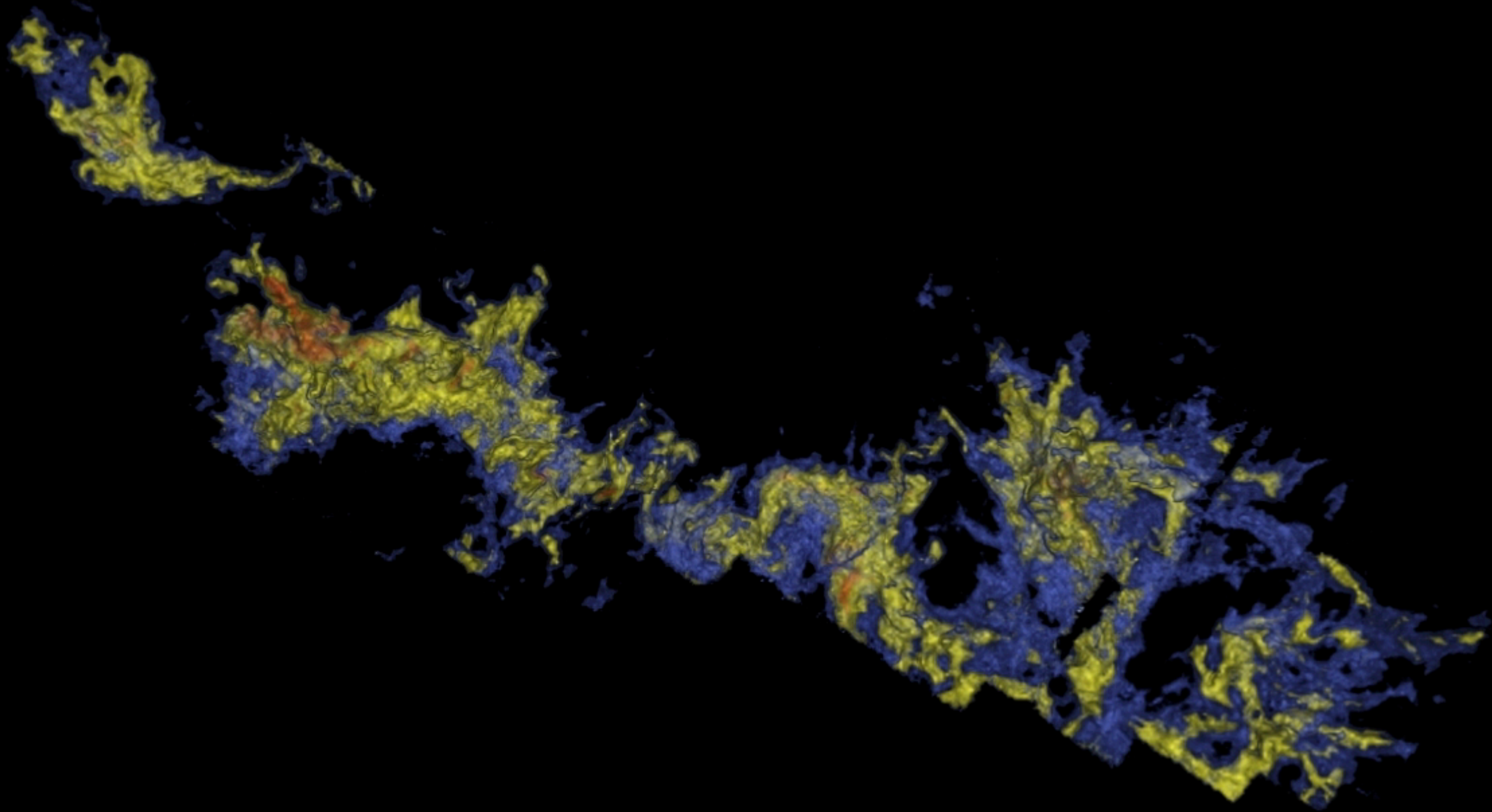

AstroMed
Perseus


Perseus



AstronomicalMedicine@IIC

COMPLETE

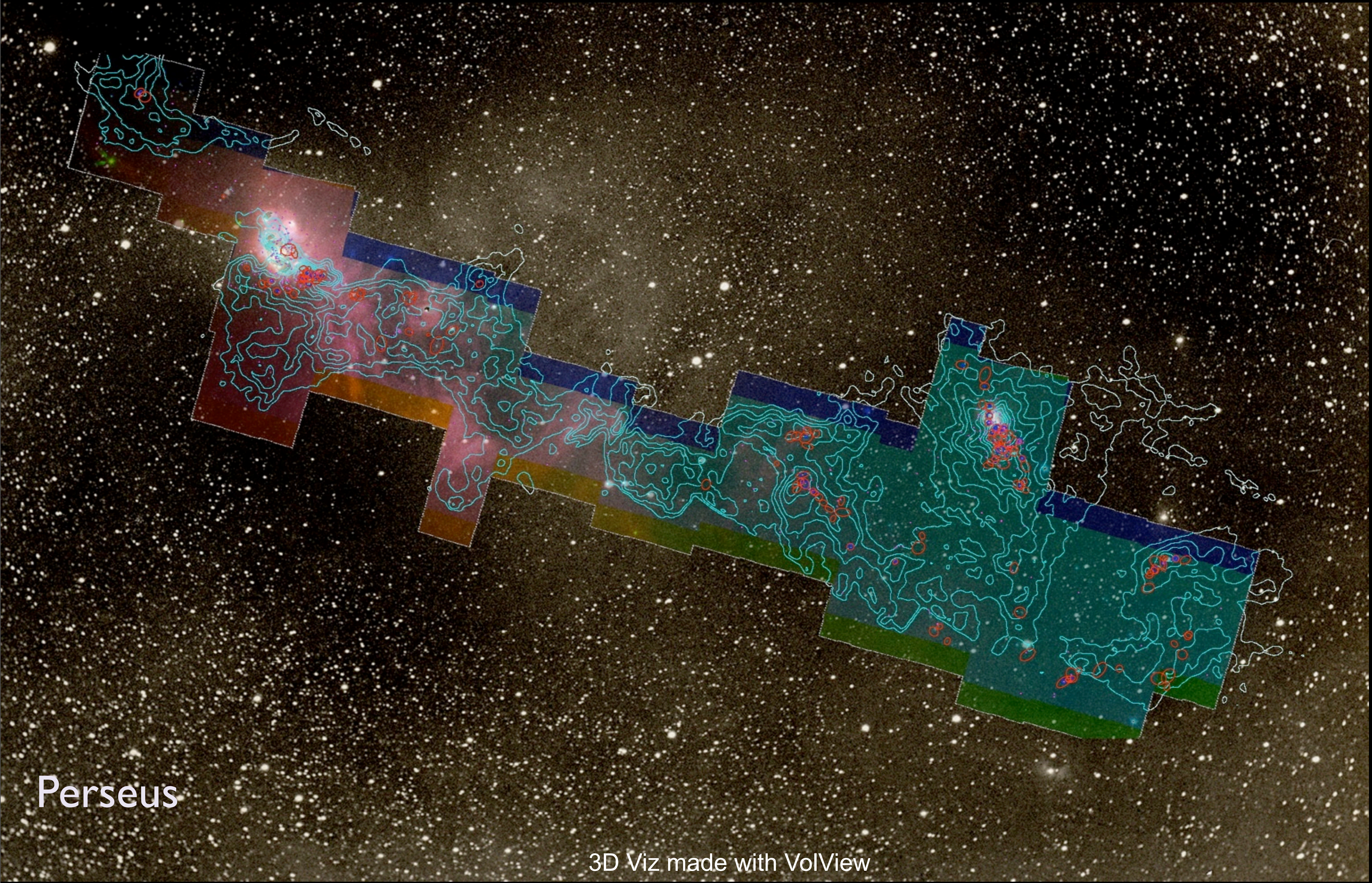


Perseus

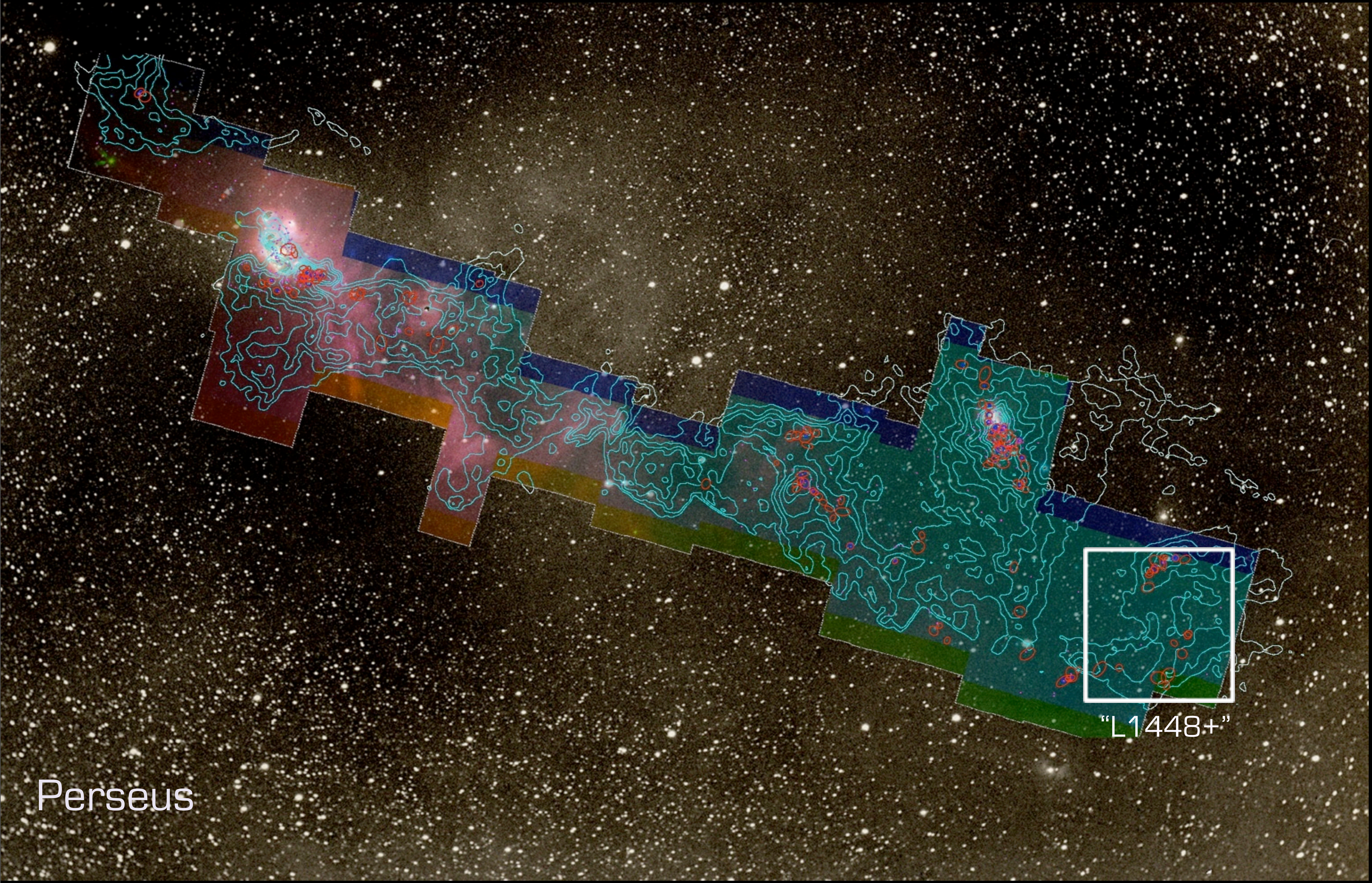
3D Viz made with VolView

AstronomicalMedicine@IIC

COMPLETE

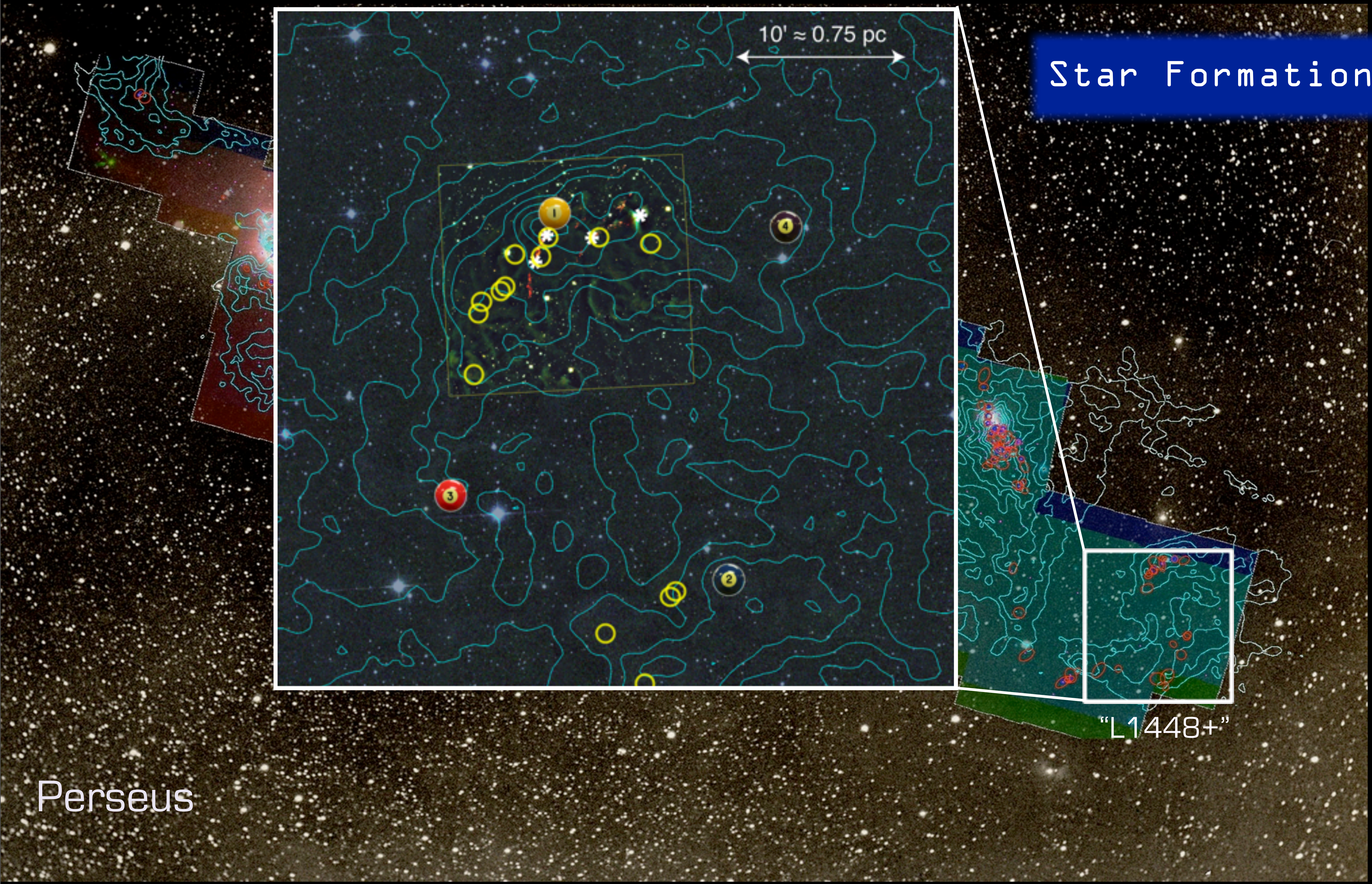


AstronomicalMedicine@IIC COMPLETE

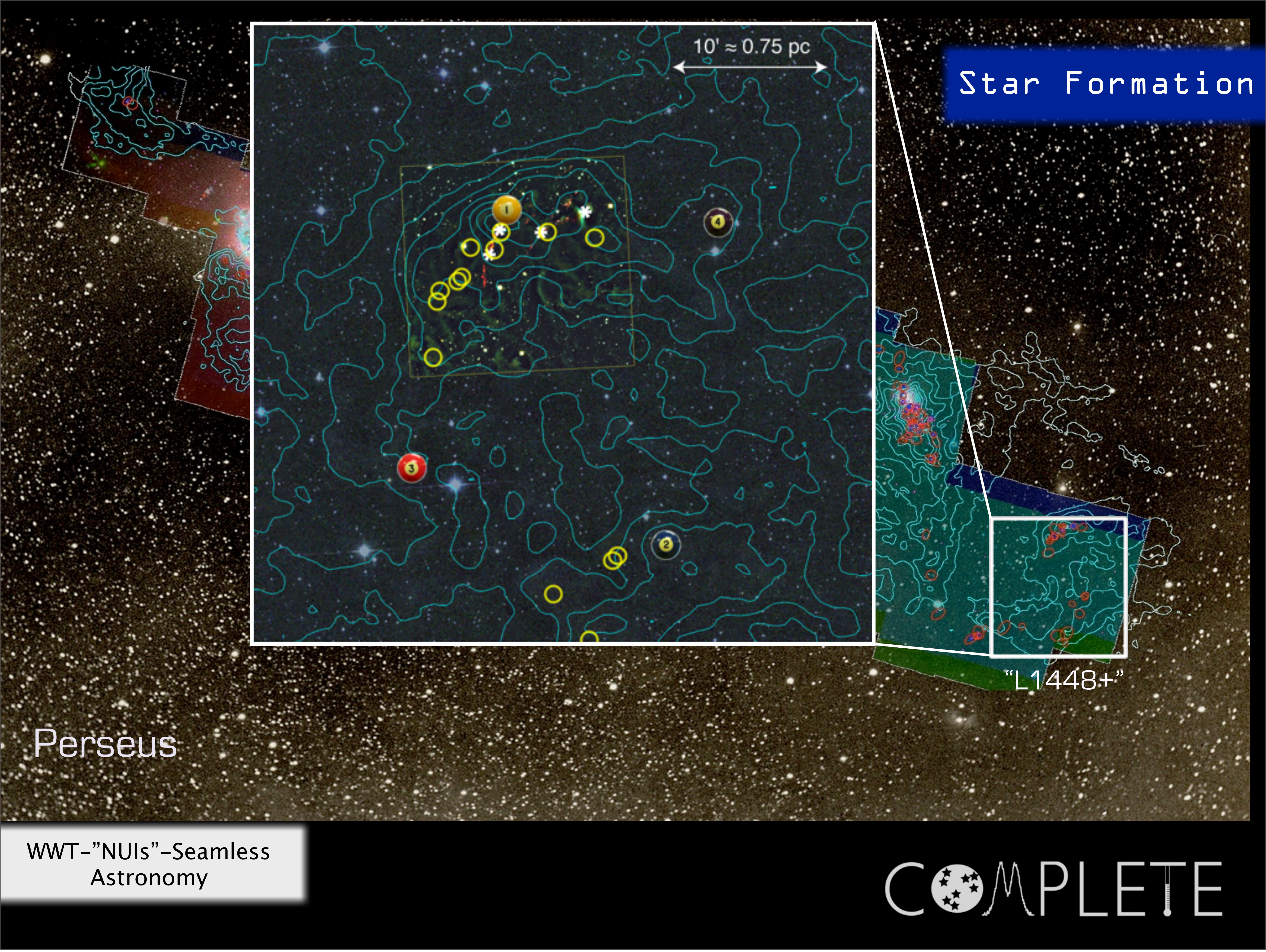


COMPLETE

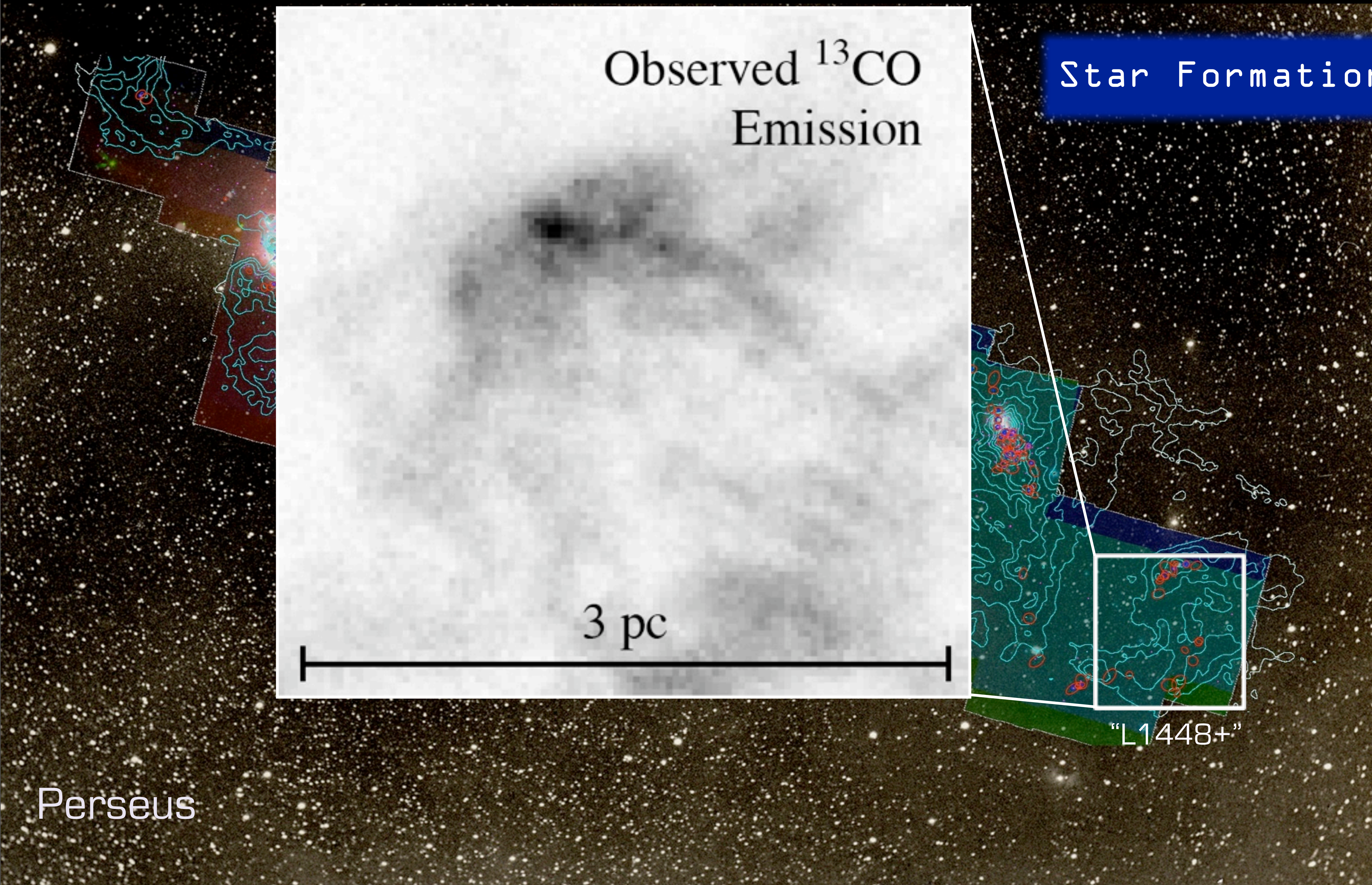
Star Formation



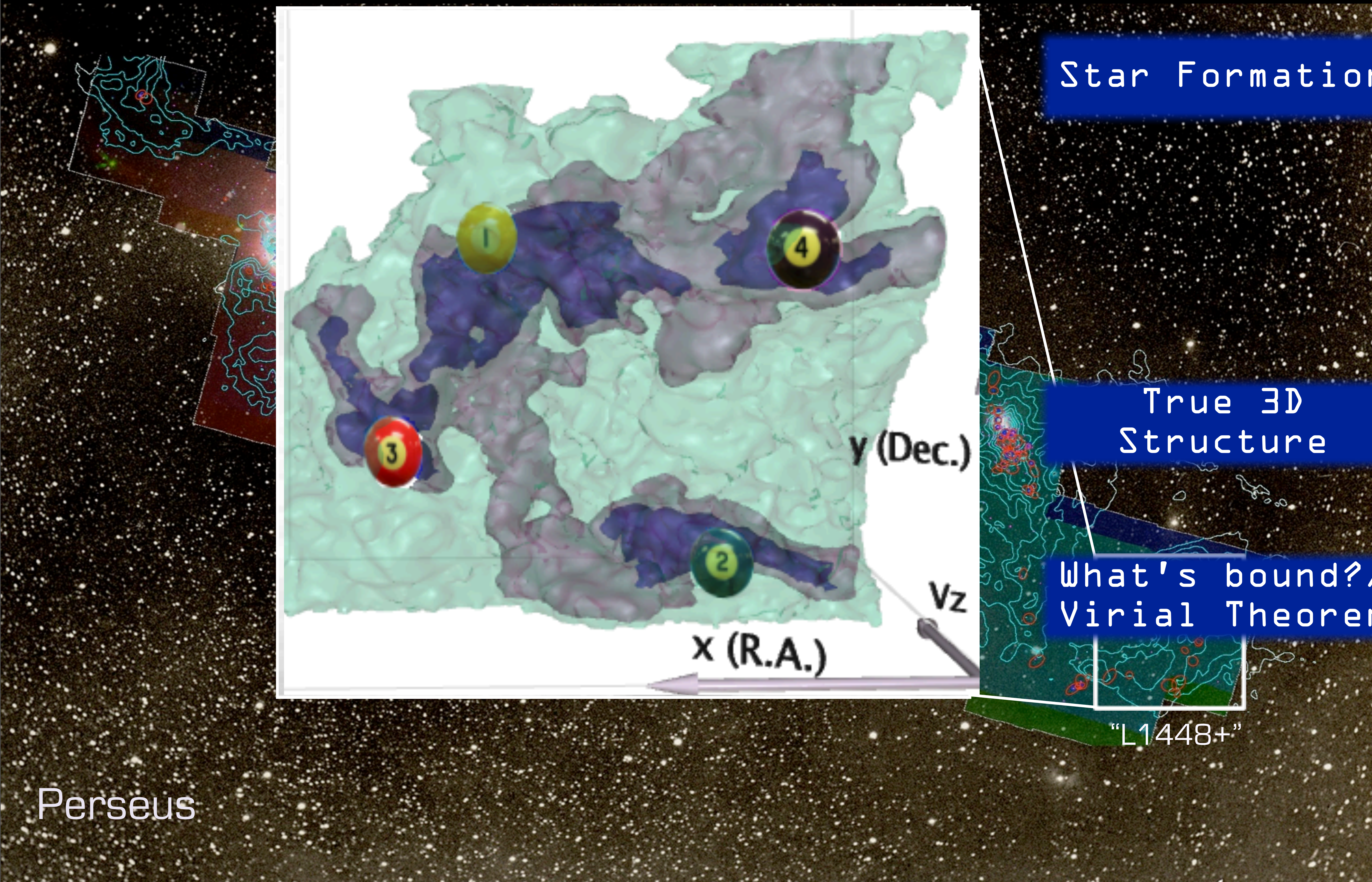
COMPLETE



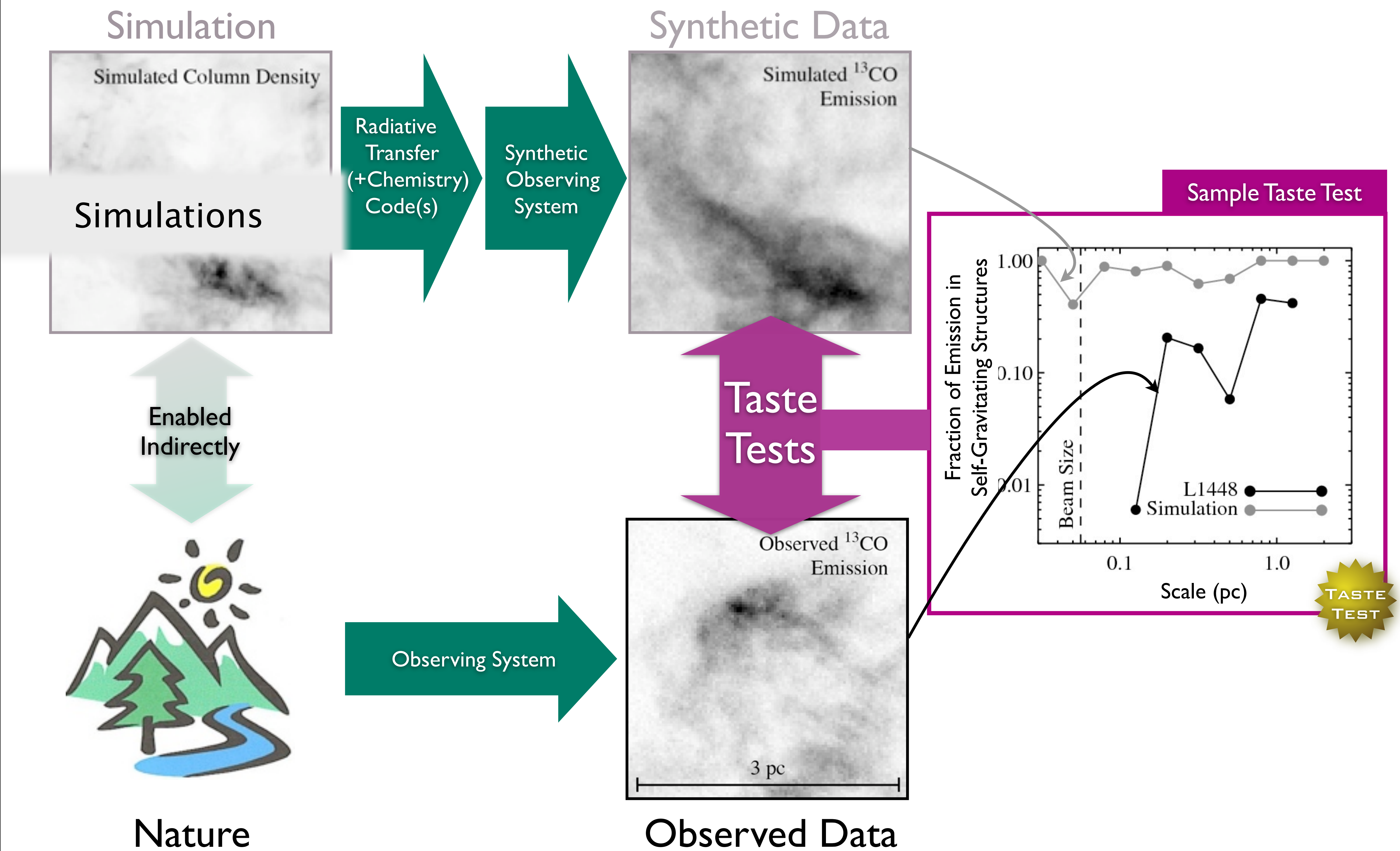
10' ≈ 0.75 pc
Star Formation
"L1448+"
Perseus
WWT-"NUIs"-Seamless
Astronomy
COMPLETE



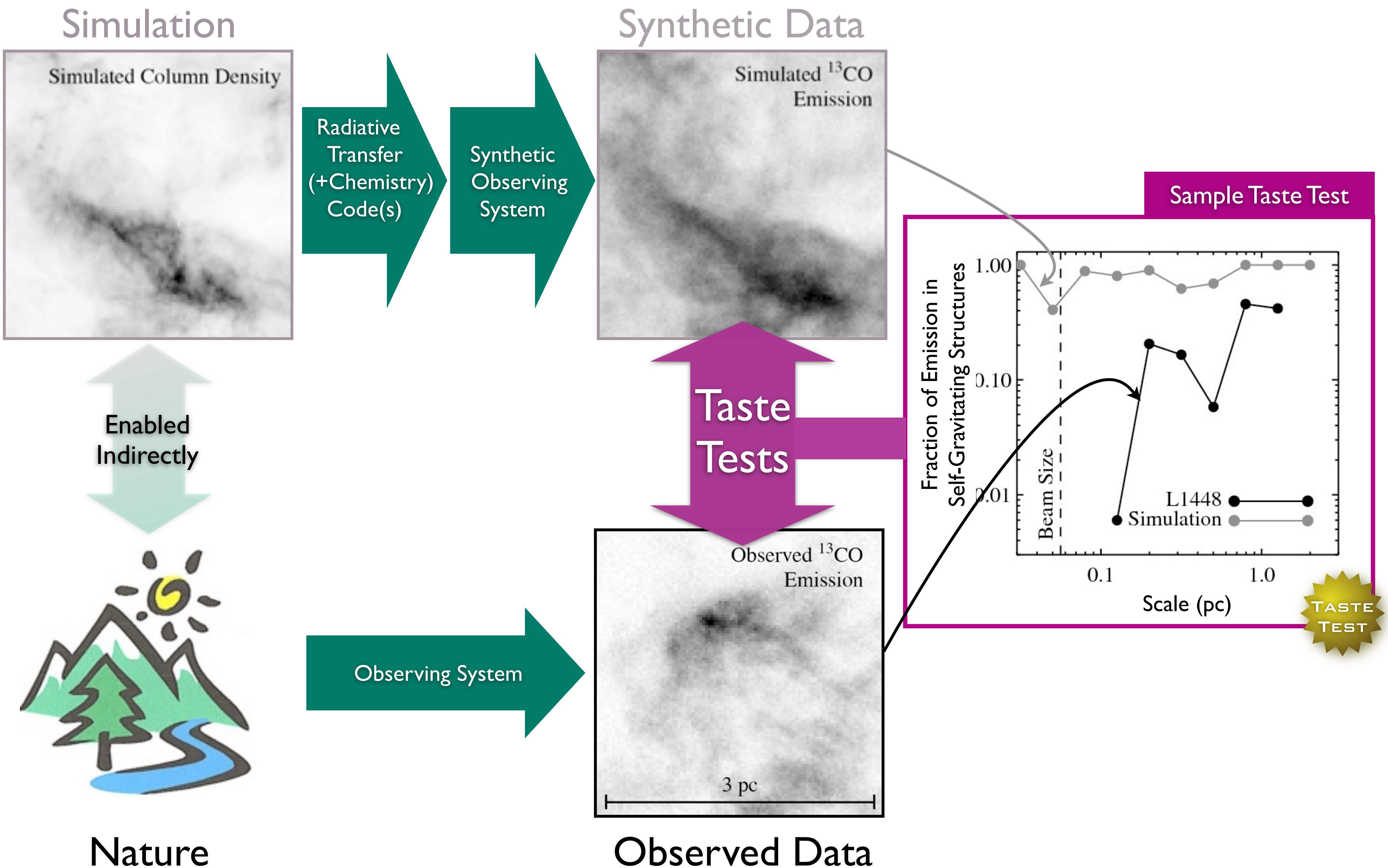
COMPLETE


Star Formation
True 3D Structure
What's bound?/
Virial Theorem
y (Dec.)
Vz
x (R.A.)
"L1448+"
Perseus
COMPLETE

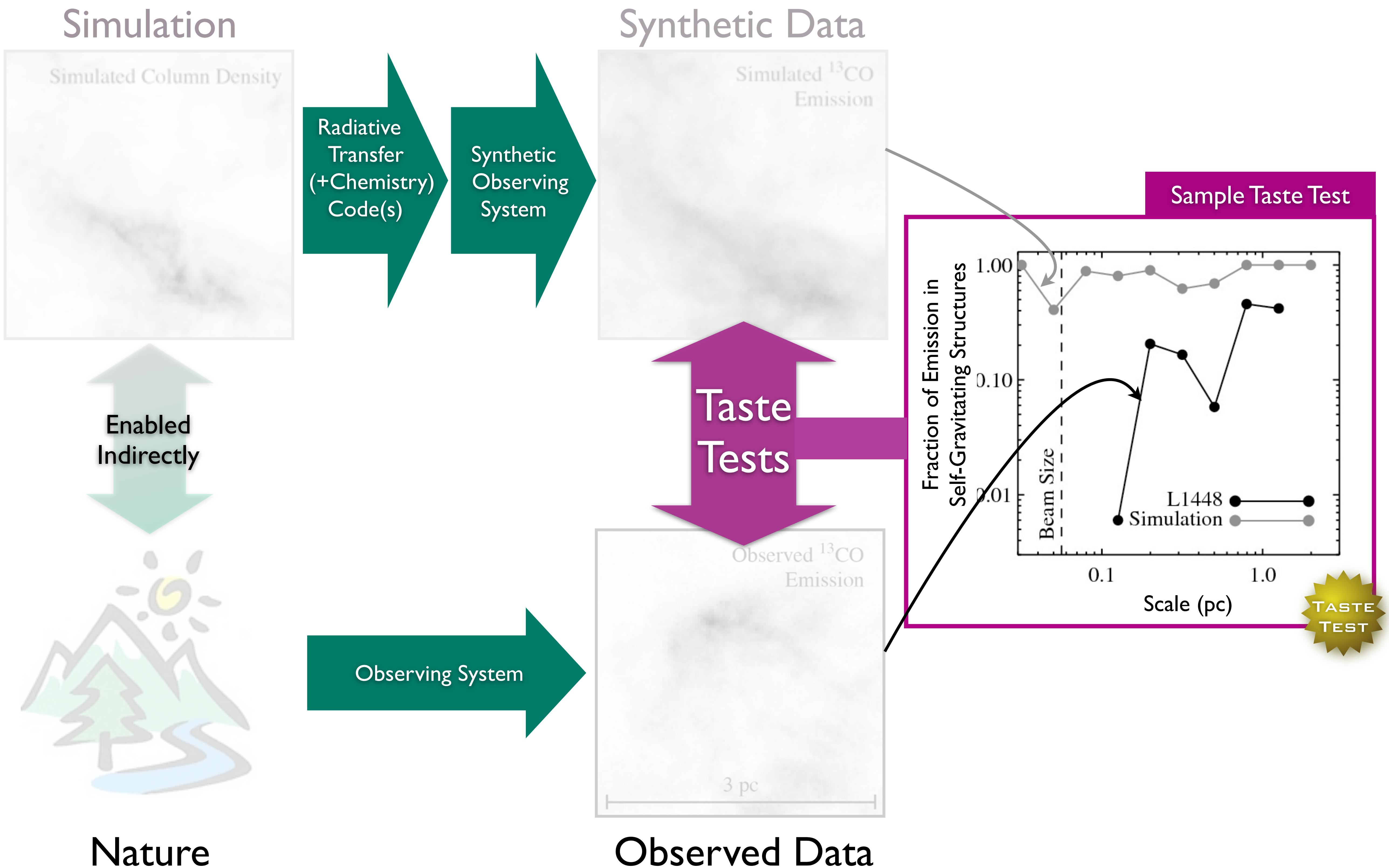
“Taste-Testing” Simulations



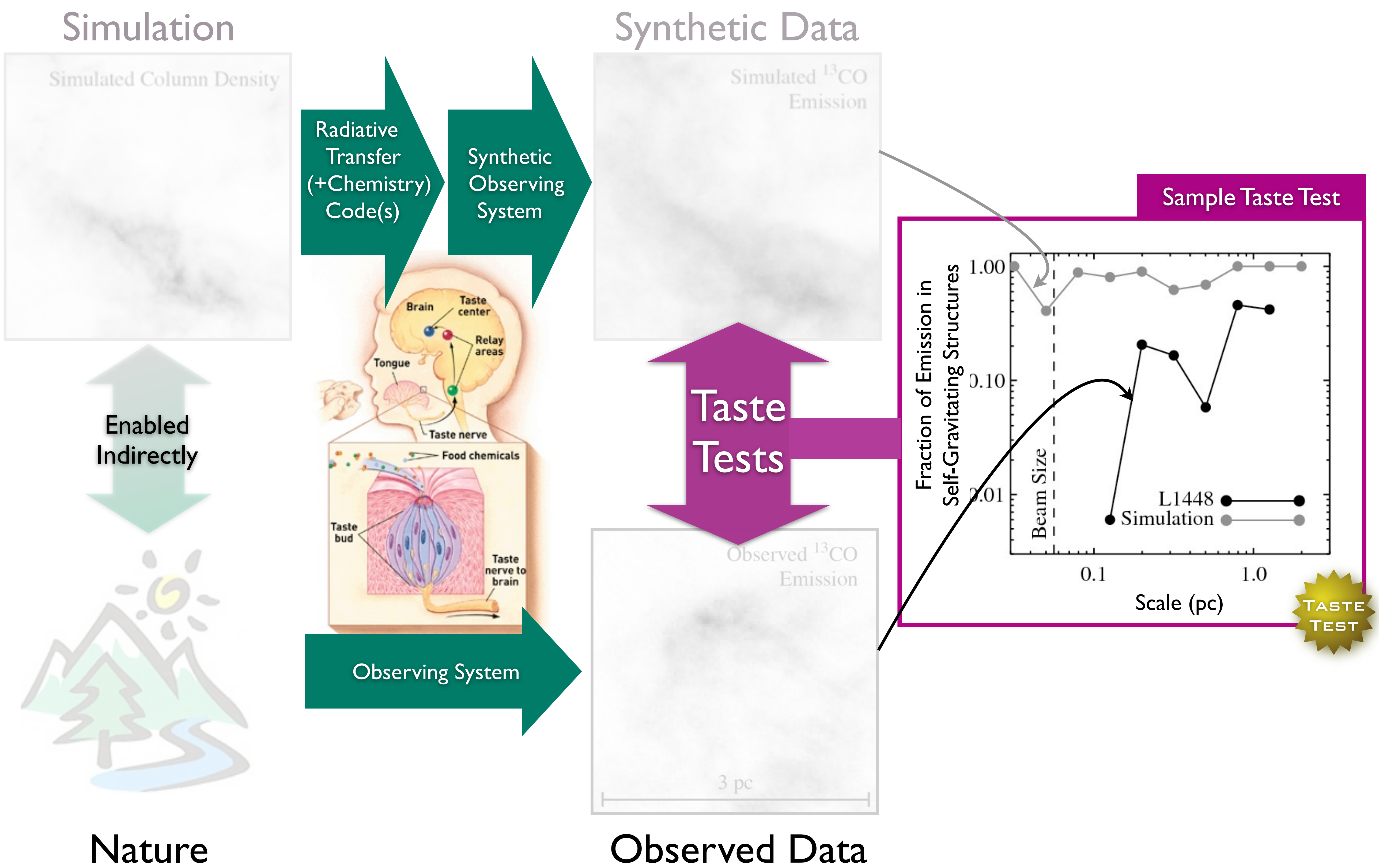
"Taste-Testing" Simulations



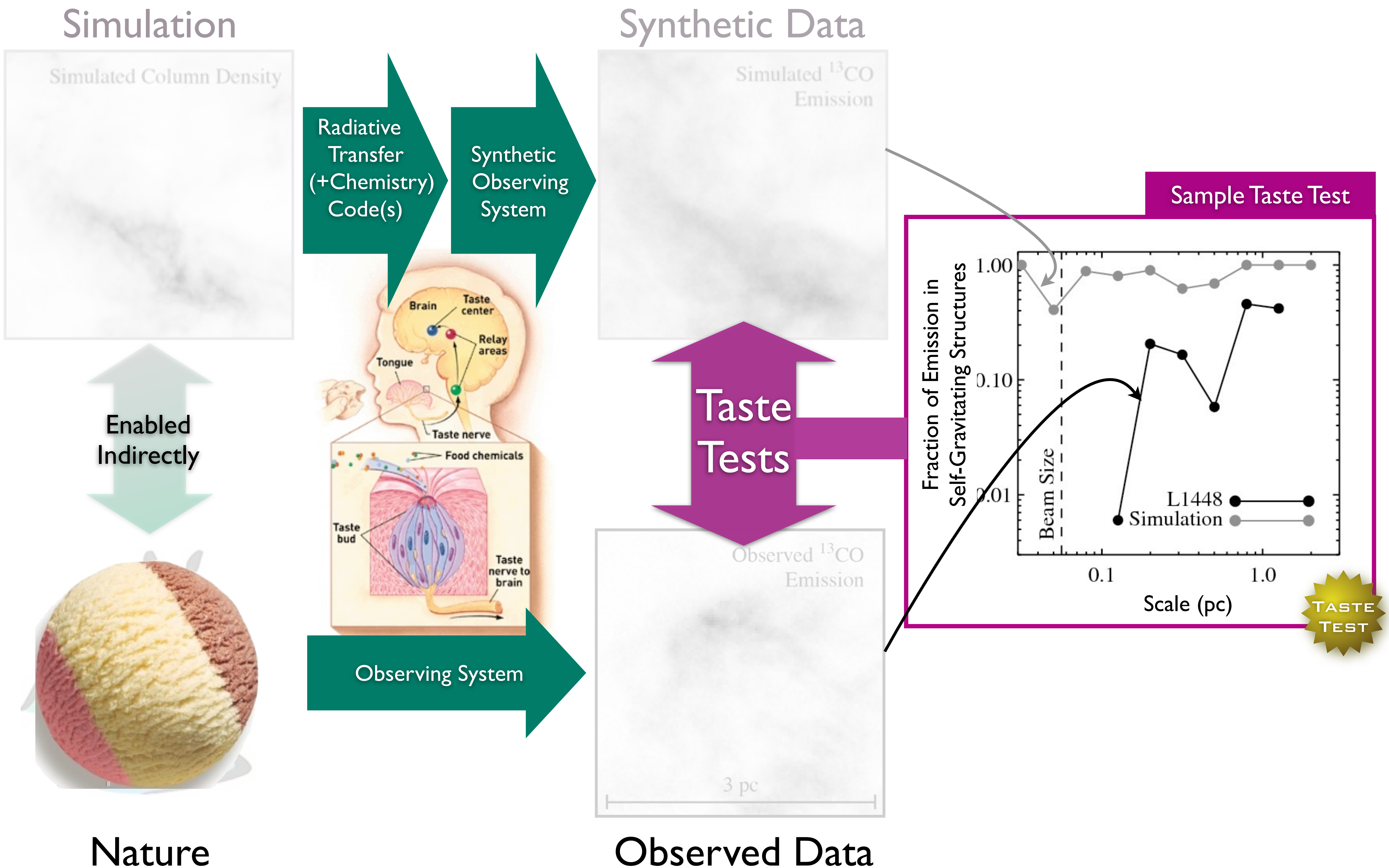
“Taste-Testing” Simulations



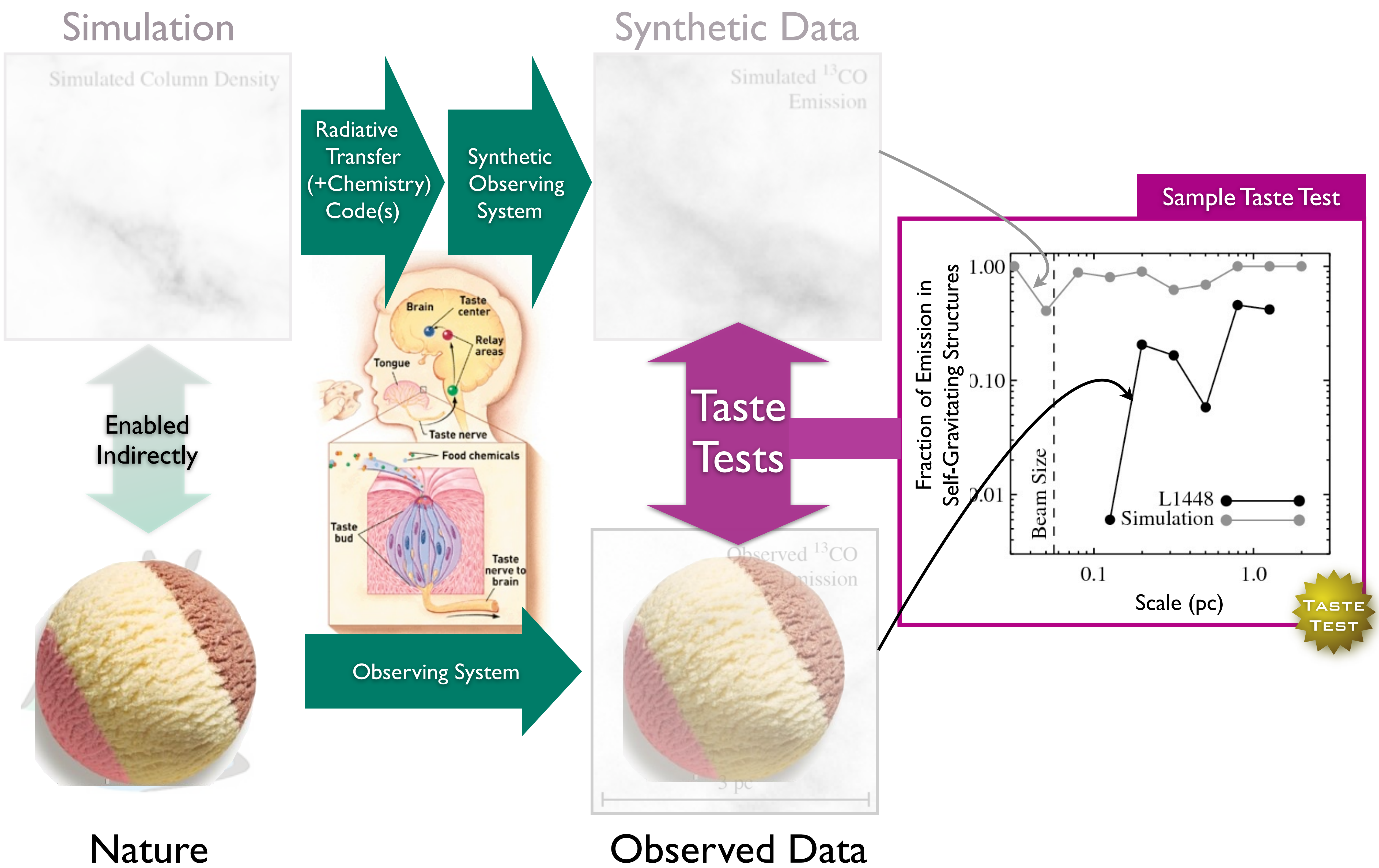
“Taste-Testing” Simulations



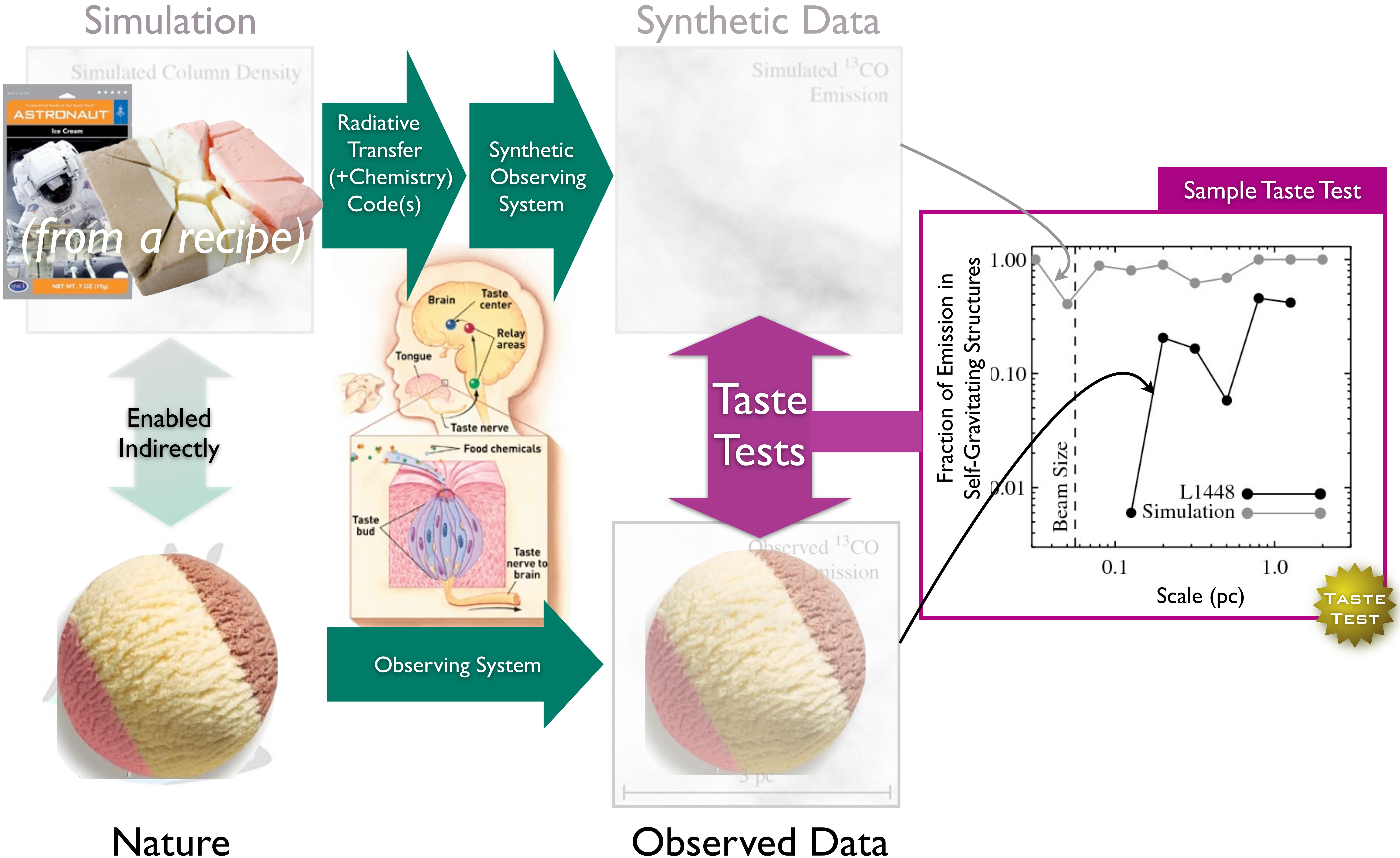
“Taste-Testing” Simulations



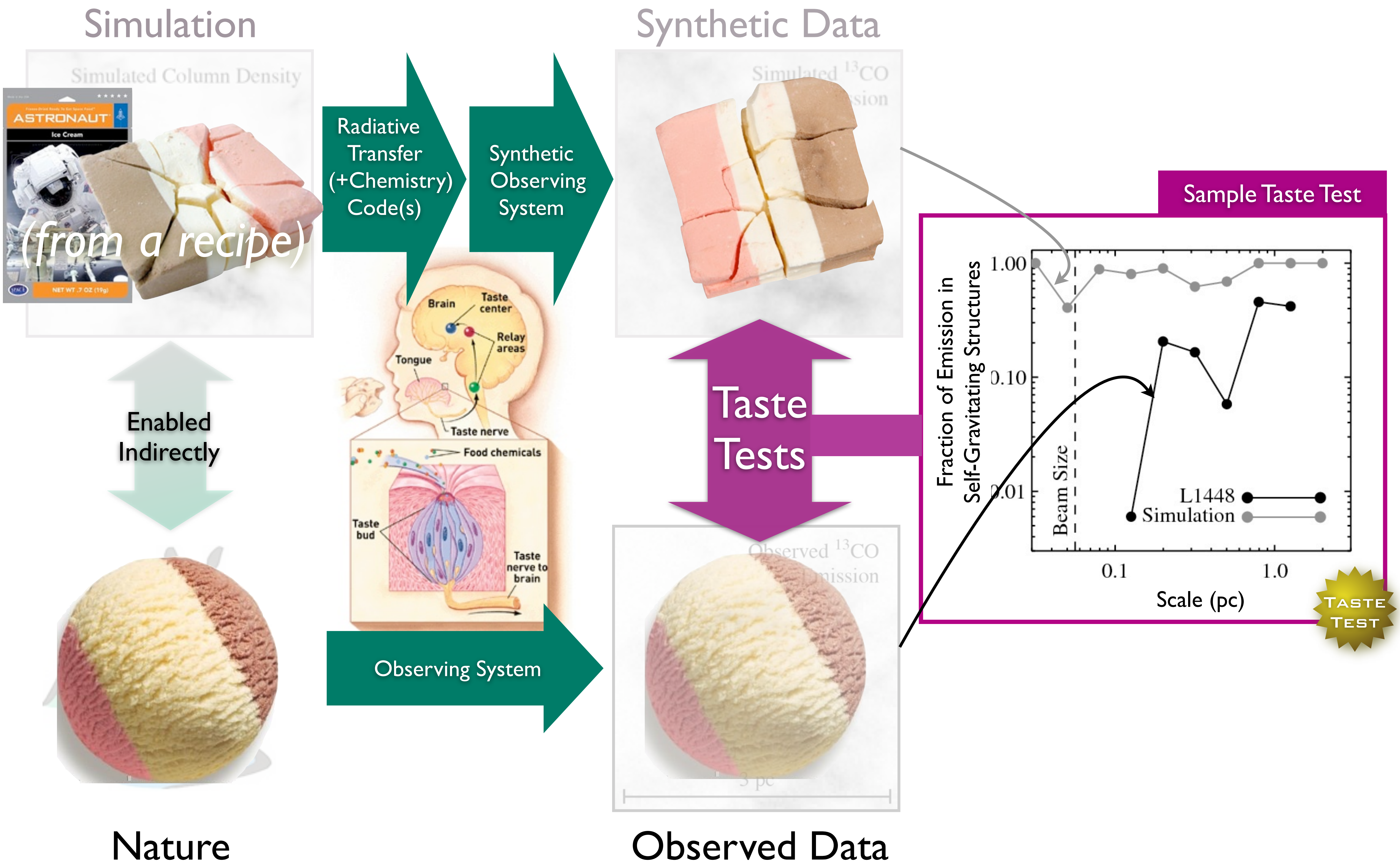
“Taste-Testing” Simulations



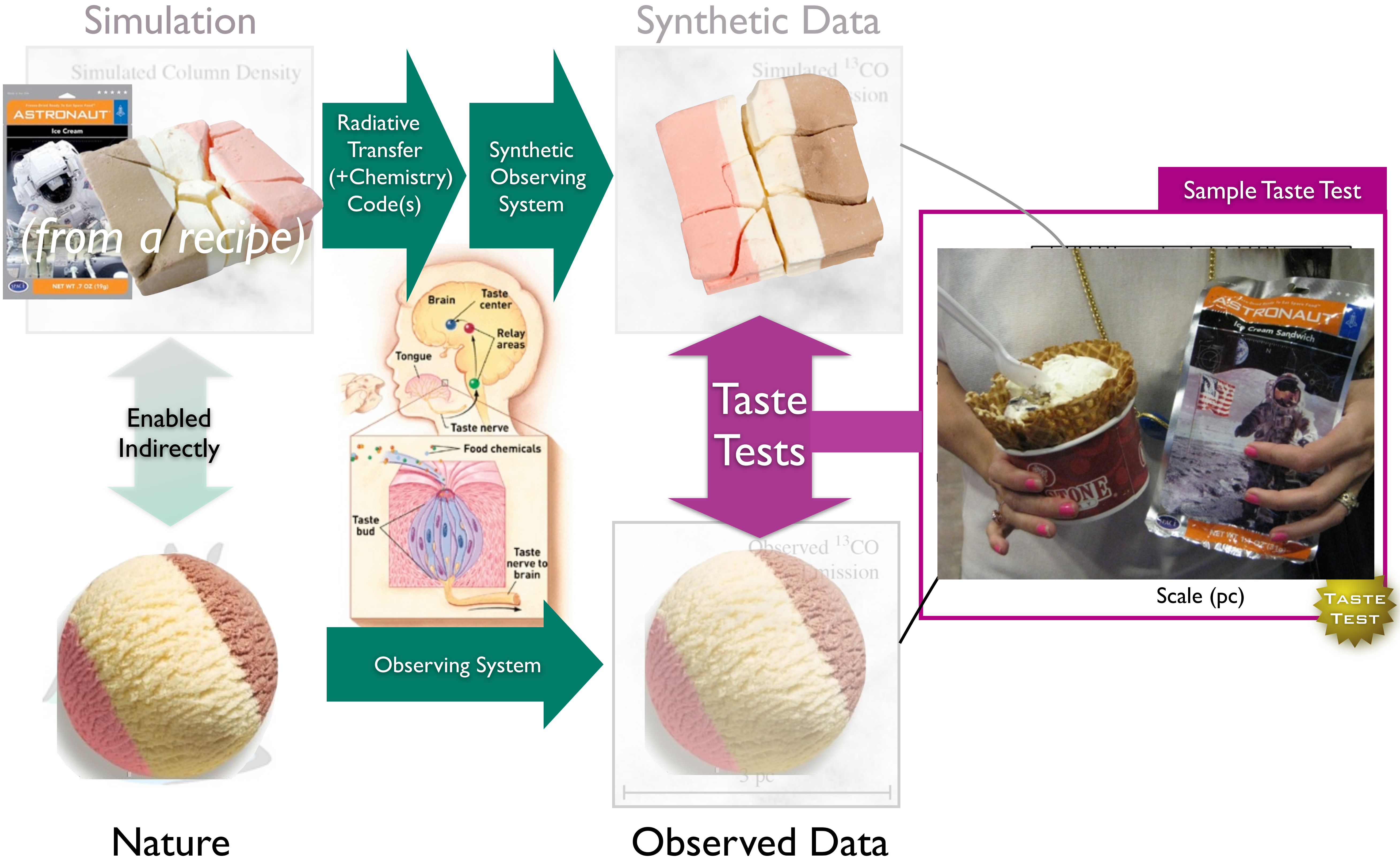
"Taste-Testing" Simulations



“Taste-Testing” Simulations



"Taste-Testing" Simulations



Taste-Testing "Gravity"

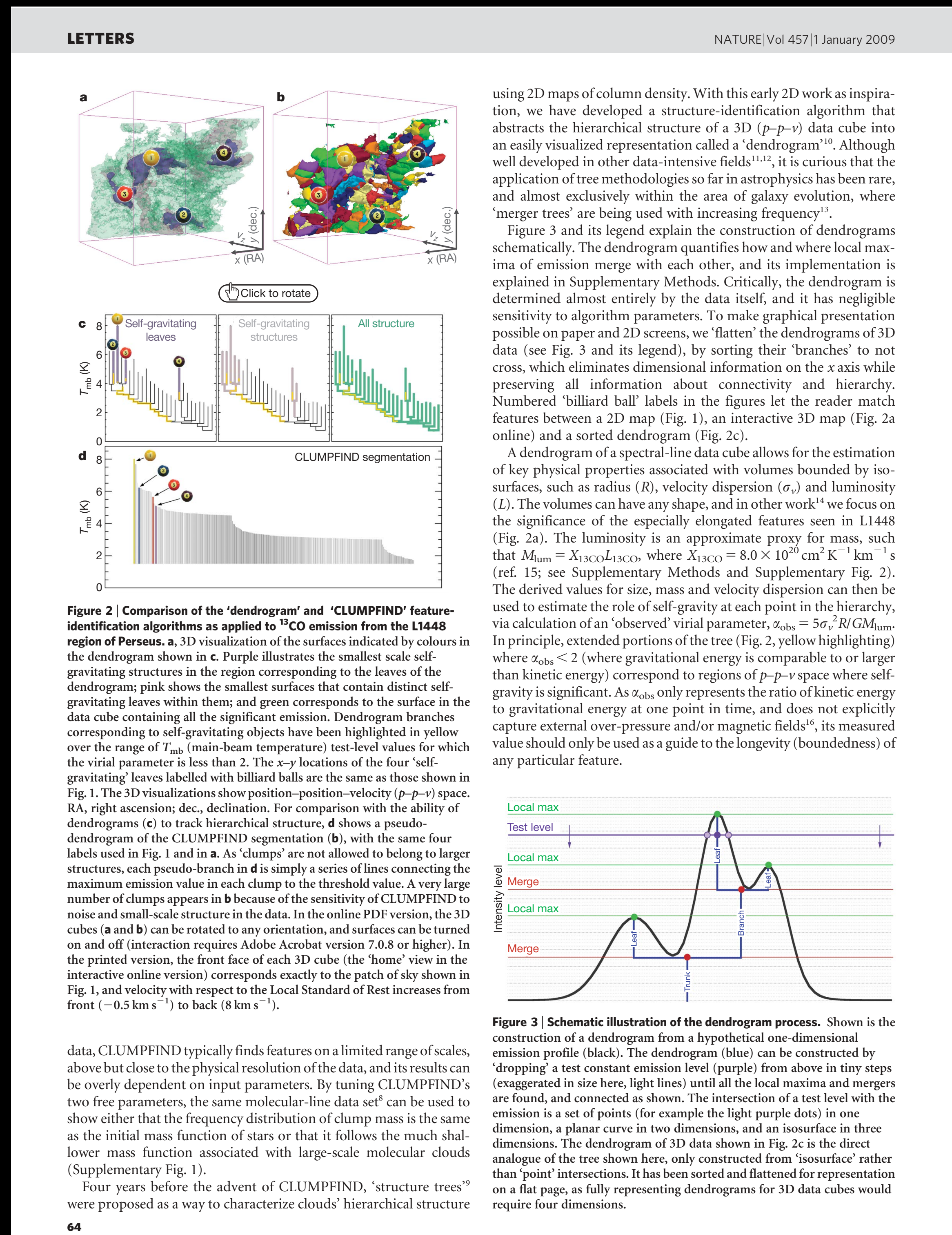


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating 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.

Goodman et al. *Nature*, 2009

Taste-Testing “Gravity”

High-Dimensional Data

AstroMed

Simulations

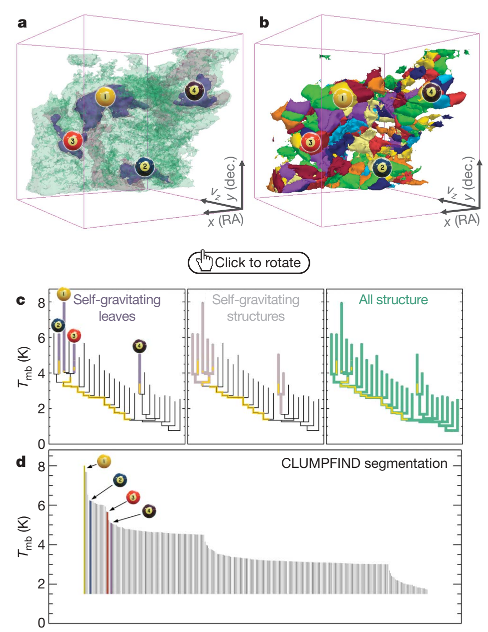


Figure 2 | Comparison of the ‘dendrogram’ and ‘CLUMPFIND’ feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating 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.

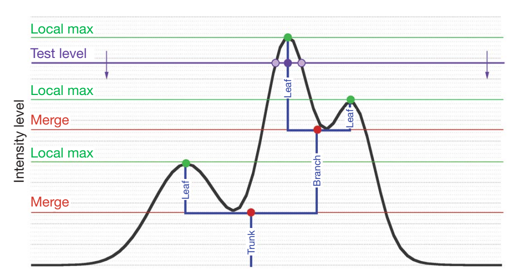


Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by ‘dropping’ a test constant emission level (purple) from above in tiny steps (exaggerated in 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.

Star Formation

“p-p-v” cubes

True 3D Structure

What's bound?/ Virial Theorem

Goodman et al. *Nature*, 2009

Taste-Testing "Gravity"

High-Dimensional Data

AstroMed

Simulations

3D PDF

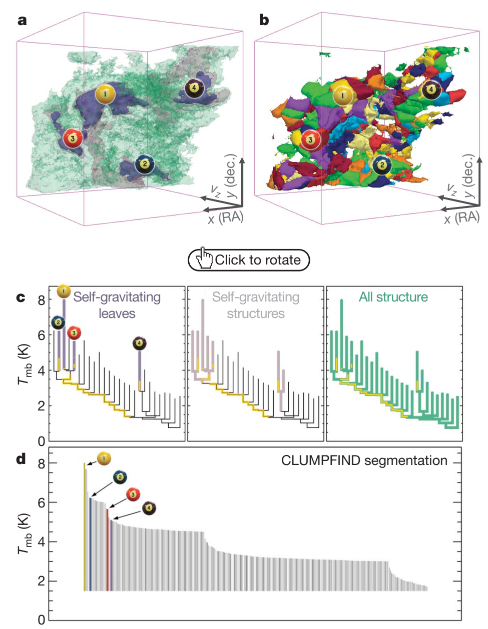


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating 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 ($-0.5\,\mathrm{km\,s^{-1}}$) to back ($8\,\mathrm{km\,s^{-1}}$).

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 $M_{\mathrm{lum}} = X_{13\mathrm{CO}} L_{13\mathrm{CO}}$, where $X_{13\mathrm{CO}} = 8.0 \times 10^{20}\,\mathrm{cm^2\,K^{-1}\,km^{-1}\,s}$ (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, $\alpha_{\mathrm{obs}} = 5\sigma_v^2 R/GM_{\mathrm{lum}}$. In principle, extended portions of the tree (Fig. 2, yellow highlighting) where $\alpha_{\mathrm{obs}} < 2$ (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.

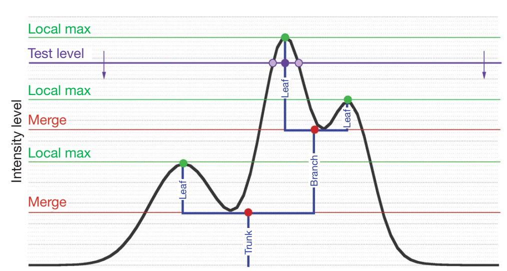


Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in 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.

Star Formation

"p-p-v" cubes

True 3D Structure

What's bound?/ Virial Theorem

Goodman et al. *Nature*, 2009

Taste-Testing "Gravity"

High-Dimensional Data

AstroMed

Simulations

3D PDF

Star Formation

"p-p-v" cubes

True 3D Structure

What's bound?/ Virial Theorem

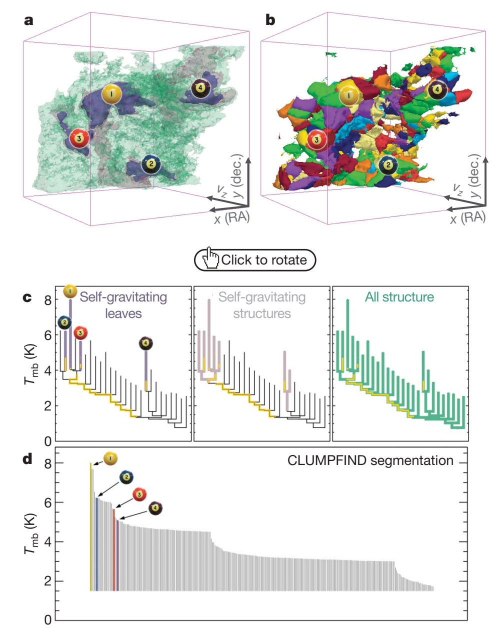


Figure 2 | Comparison of the 'dendrogram' and 'CLUMPFIND' feature-identification algorithms as applied to ^{13}CO emission from the L1448 region of Perseus. a, 3D visualization of the surfaces indicated by colours in the dendrogram shown in **c**. Purple illustrates the smallest scale self-gravitating 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) s[…]ace. RA, right ascension; dec., declination. For comparison with the ability [...] dendrograms (**c**) to track hierarchical structure, **d** shows a pseudo-dendrogram of the CLUMPFIND segmentation (**b**), with the same fou[...] [...] Fig. 1, and velocity with respect to the Local Standard of Rest increases [...] front ($-0.5\ \mathrm{km\,s^{-1}}$) to back ($8\ \mathrm{km\,s^{-1}}$).

"Dendrogram"

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 $X_{13CO} = 8.0 \times 10^{20}\ \mathrm{cm^2\,K^{-1}\,km^{-1}\,s}$ (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.

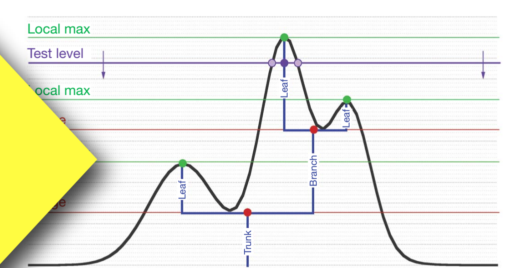
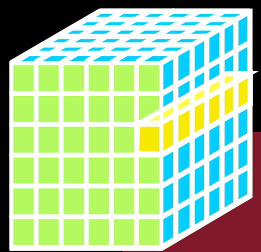


Figure 3 | Schematic illustration of the dendrogram process. Shown is the construction of a dendrogram from a hypothetical one-dimensional emission profile (black). The dendrogram (blue) can be constructed by 'dropping' a test constant emission level (purple) from above in tiny steps (exaggerated in 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.

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).

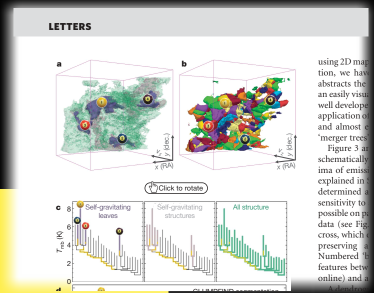
Four years before the advent of CLUMPFIND, 'structure trees'[9] were proposed as a way to characterize clouds' hierarchical structure

Goodman et al. *Nature*, 2009



Data Dimensions Display

Linked Views

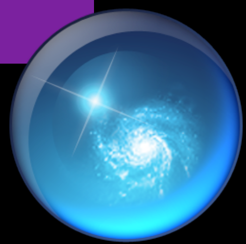


"Taste-Testing"
[Sim:Data]

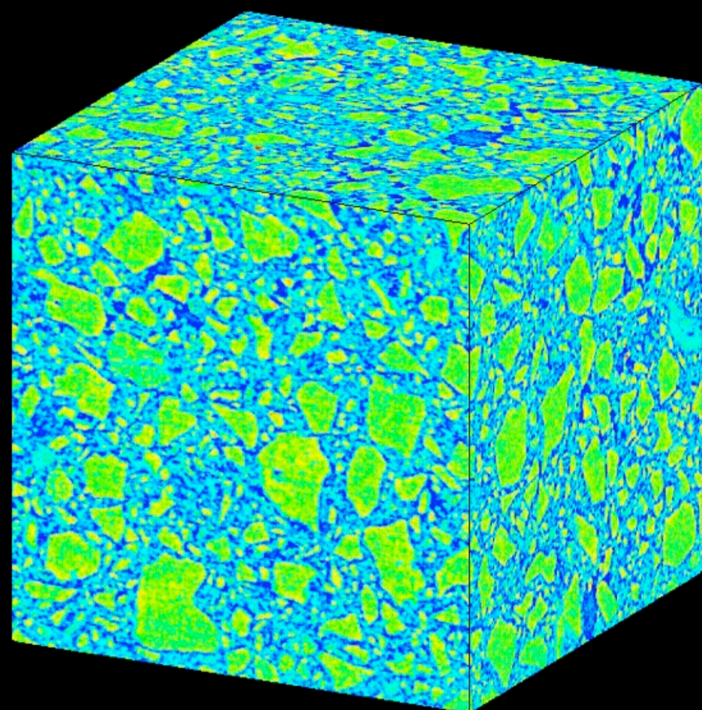
Seamless Astronomy
[the future]

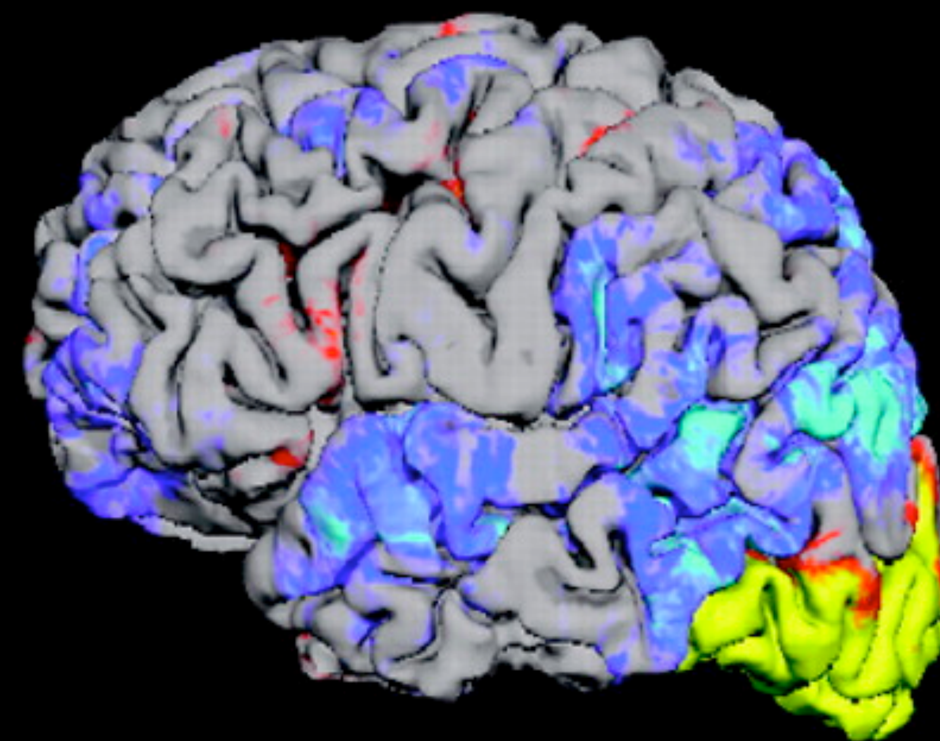
ADS Labs
[info viz]

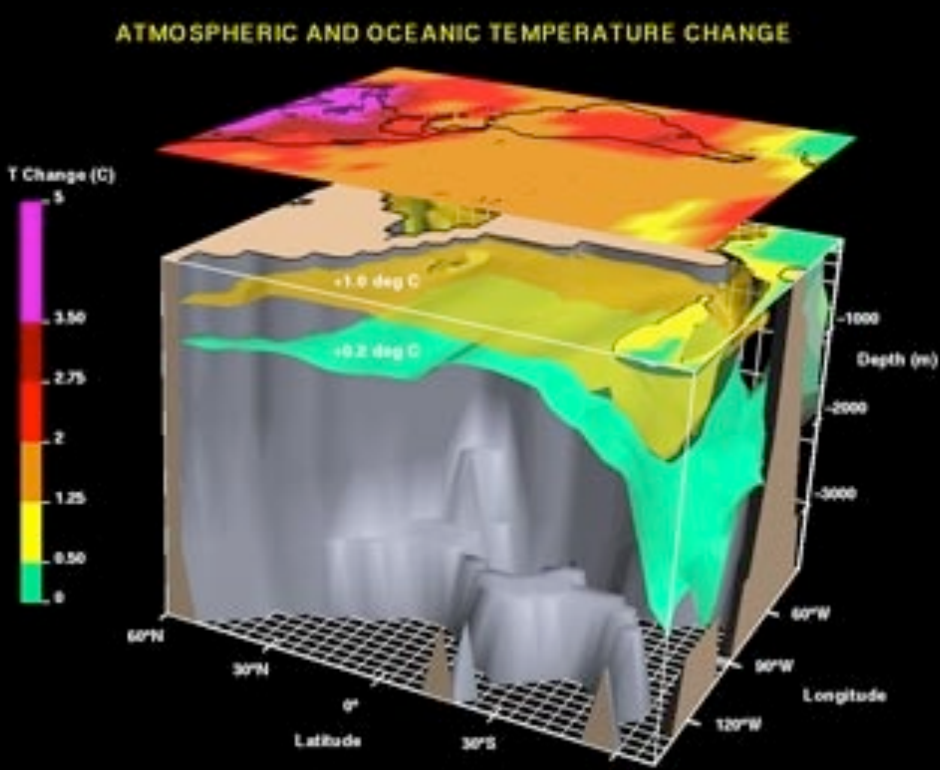
Open Data
Open Tools

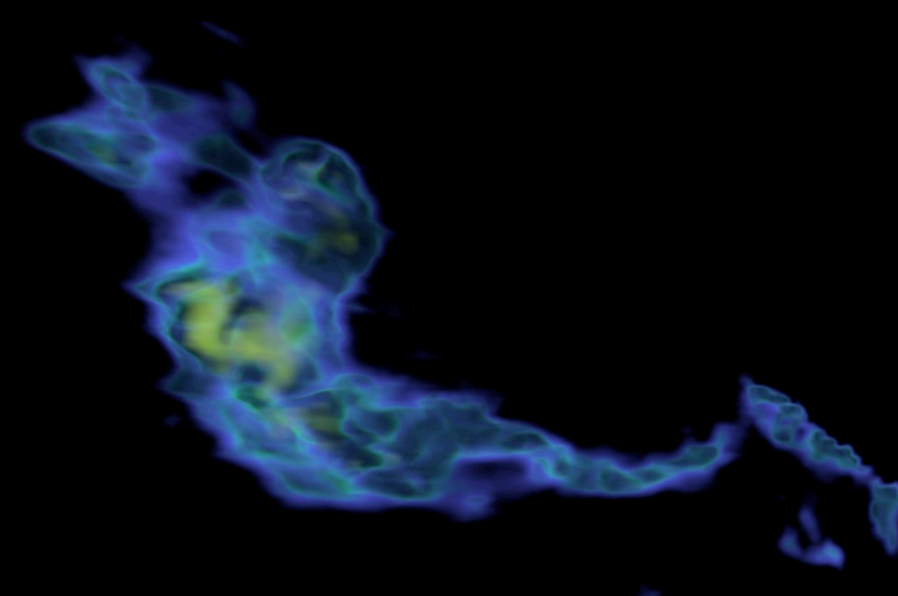


“High-Dimensional” Data

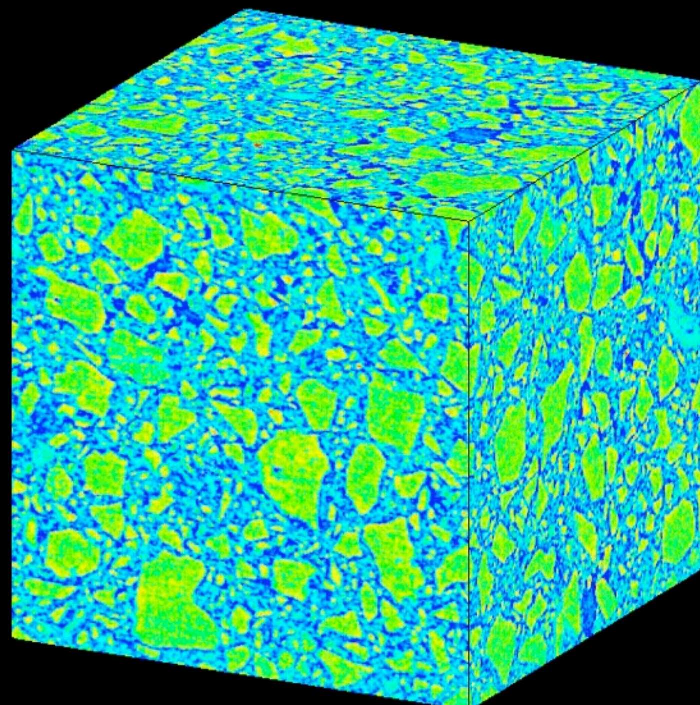


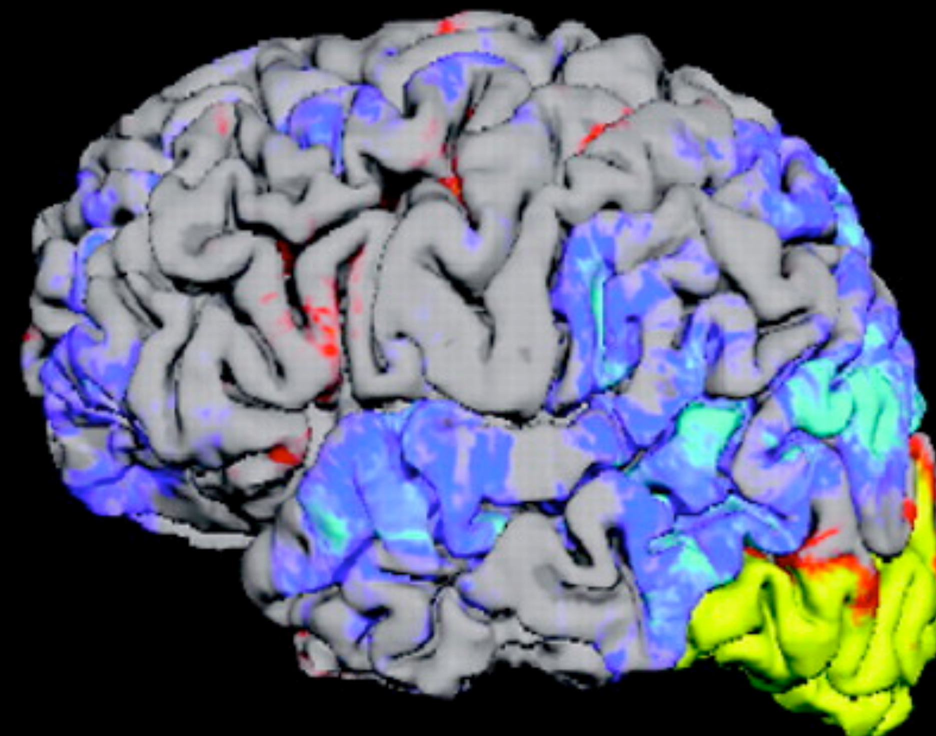


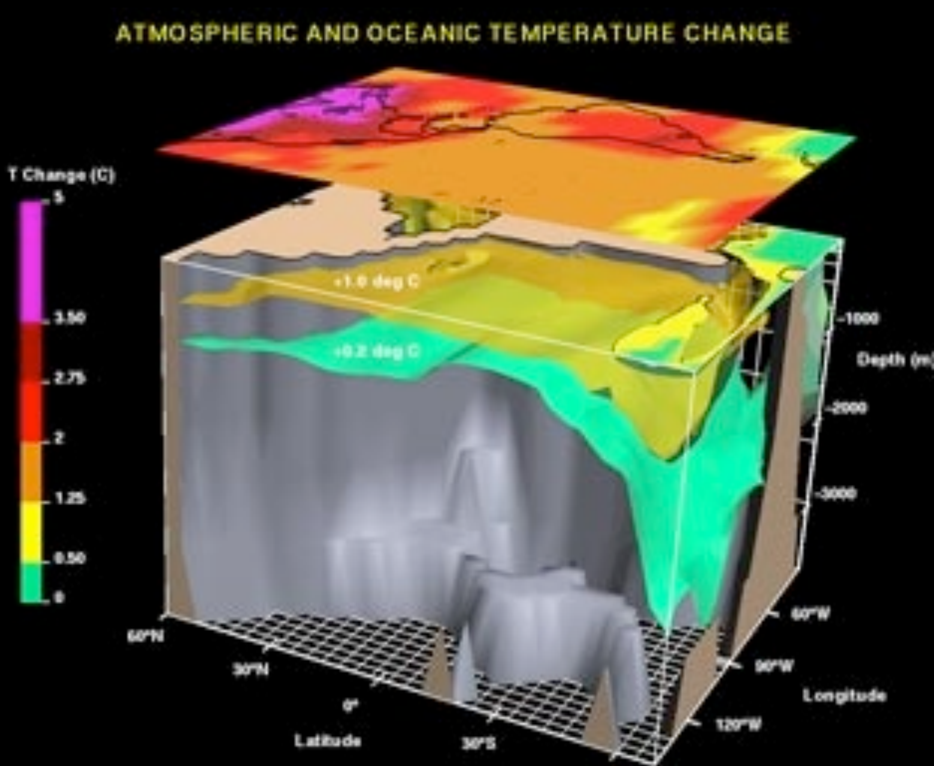


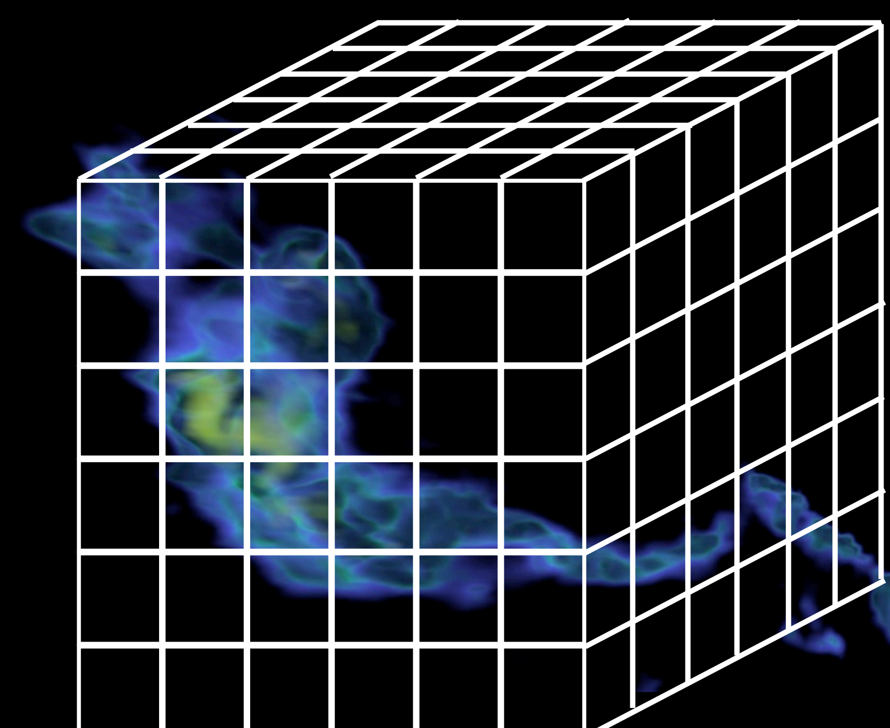


“High-Dimensional” Data

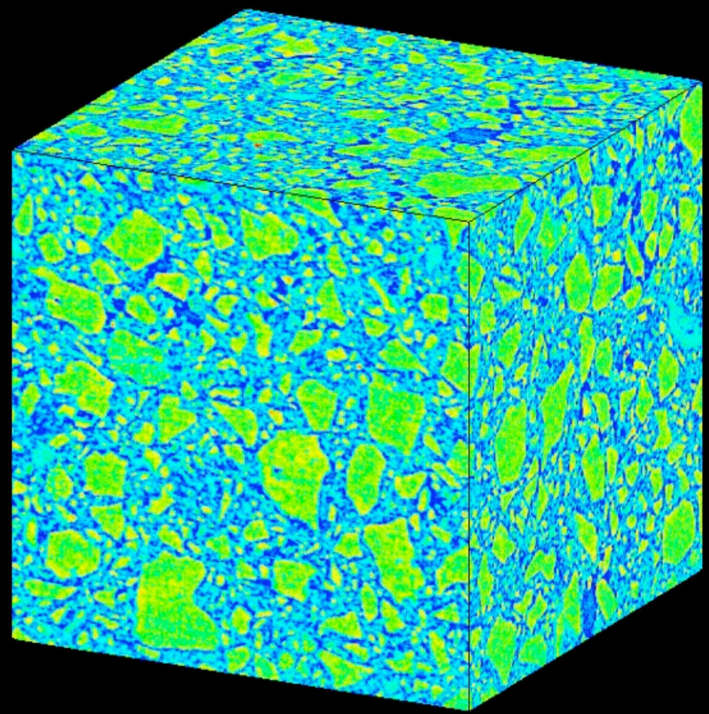


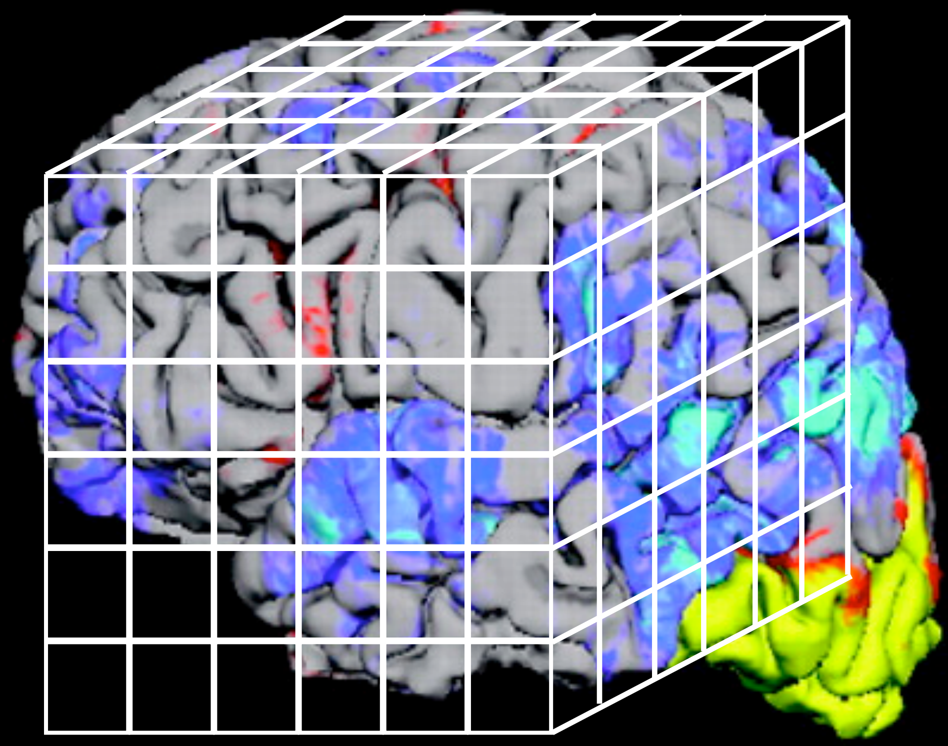


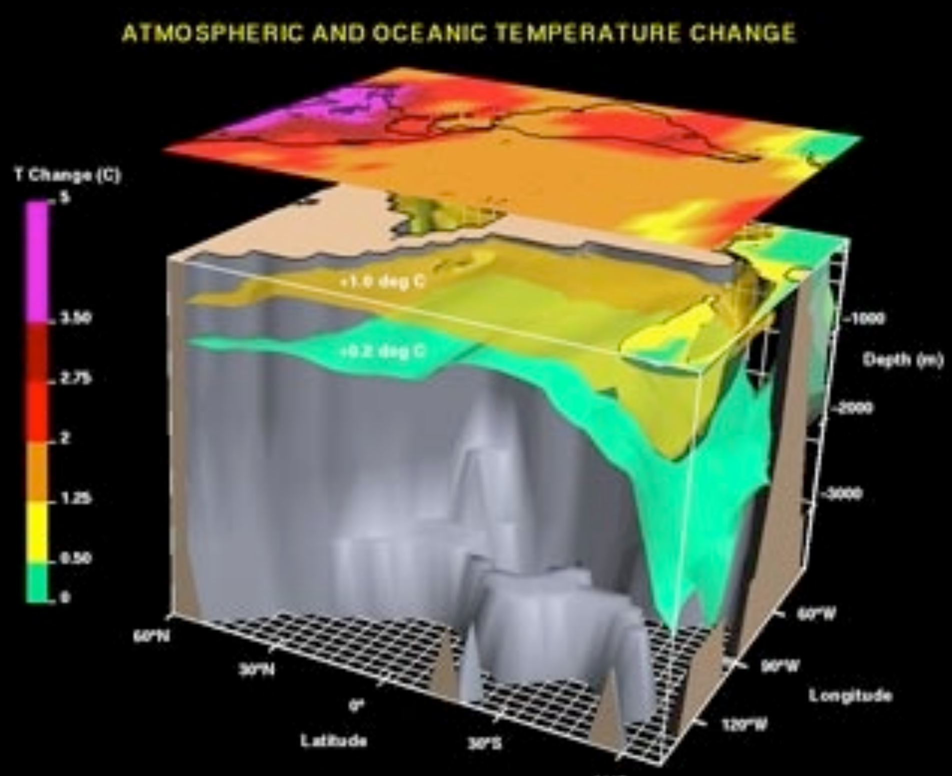


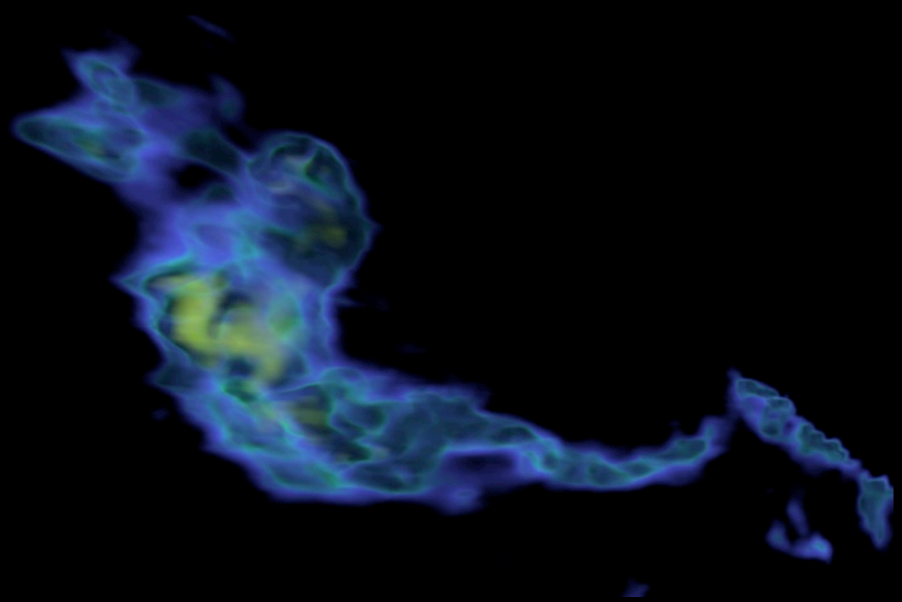


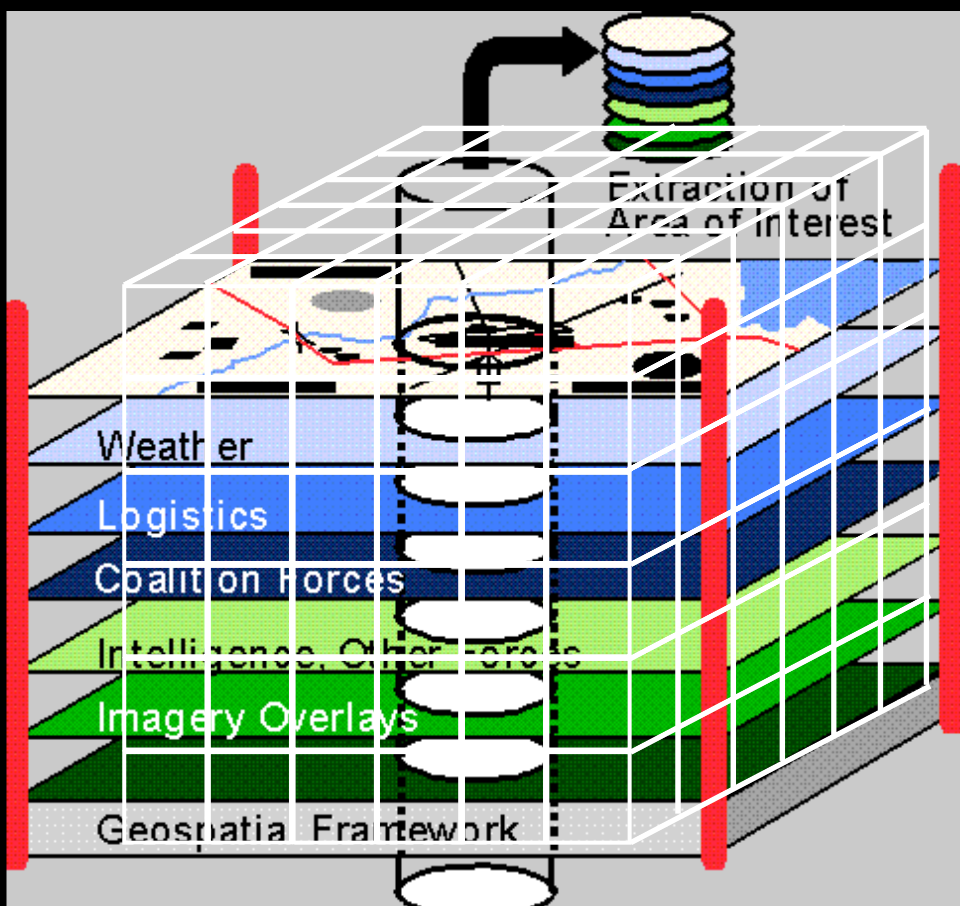
“High-Dimensional” Data



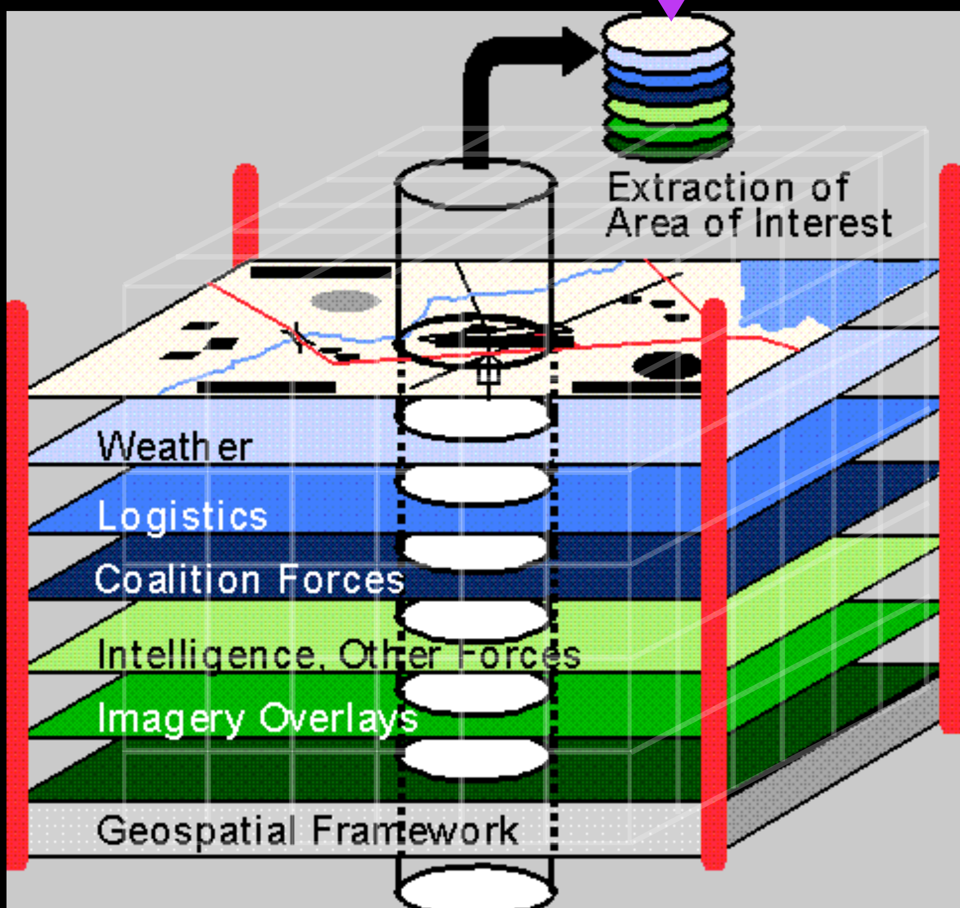






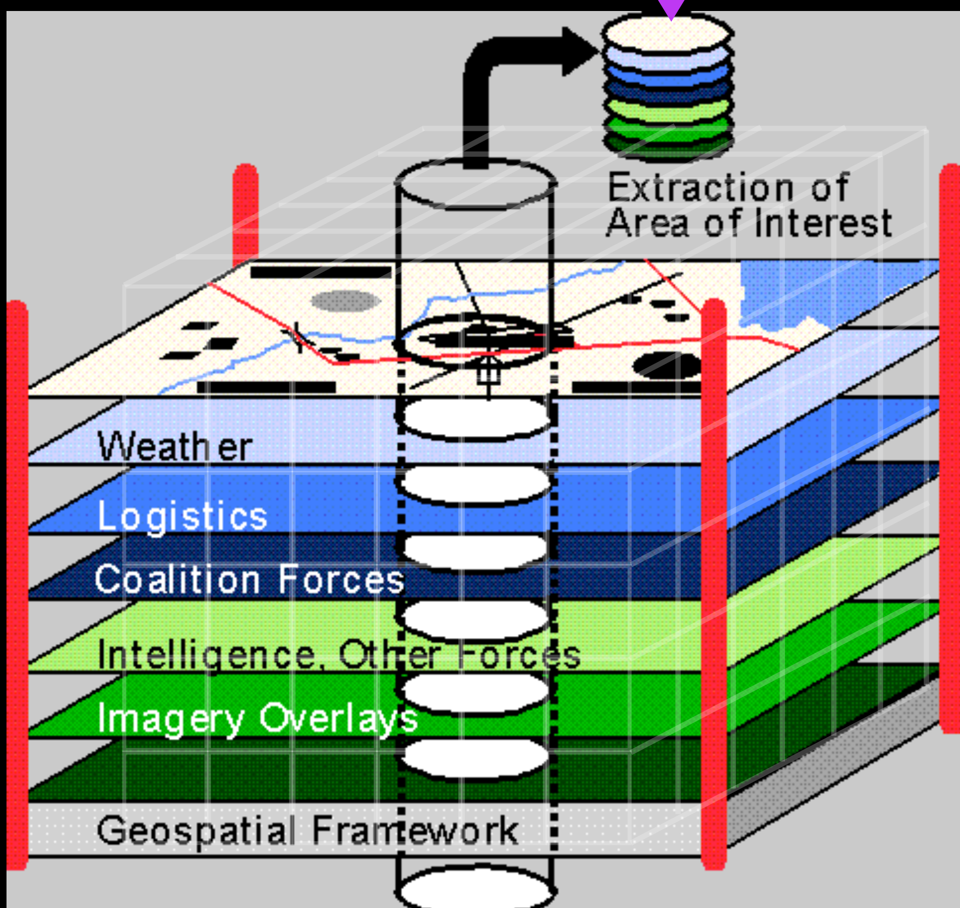

Extraction of
Area of Interest
Weather
Logistics
Coalition Forces
Intelligence, Other Forces
Imagery Overlays
Geospatial Framework

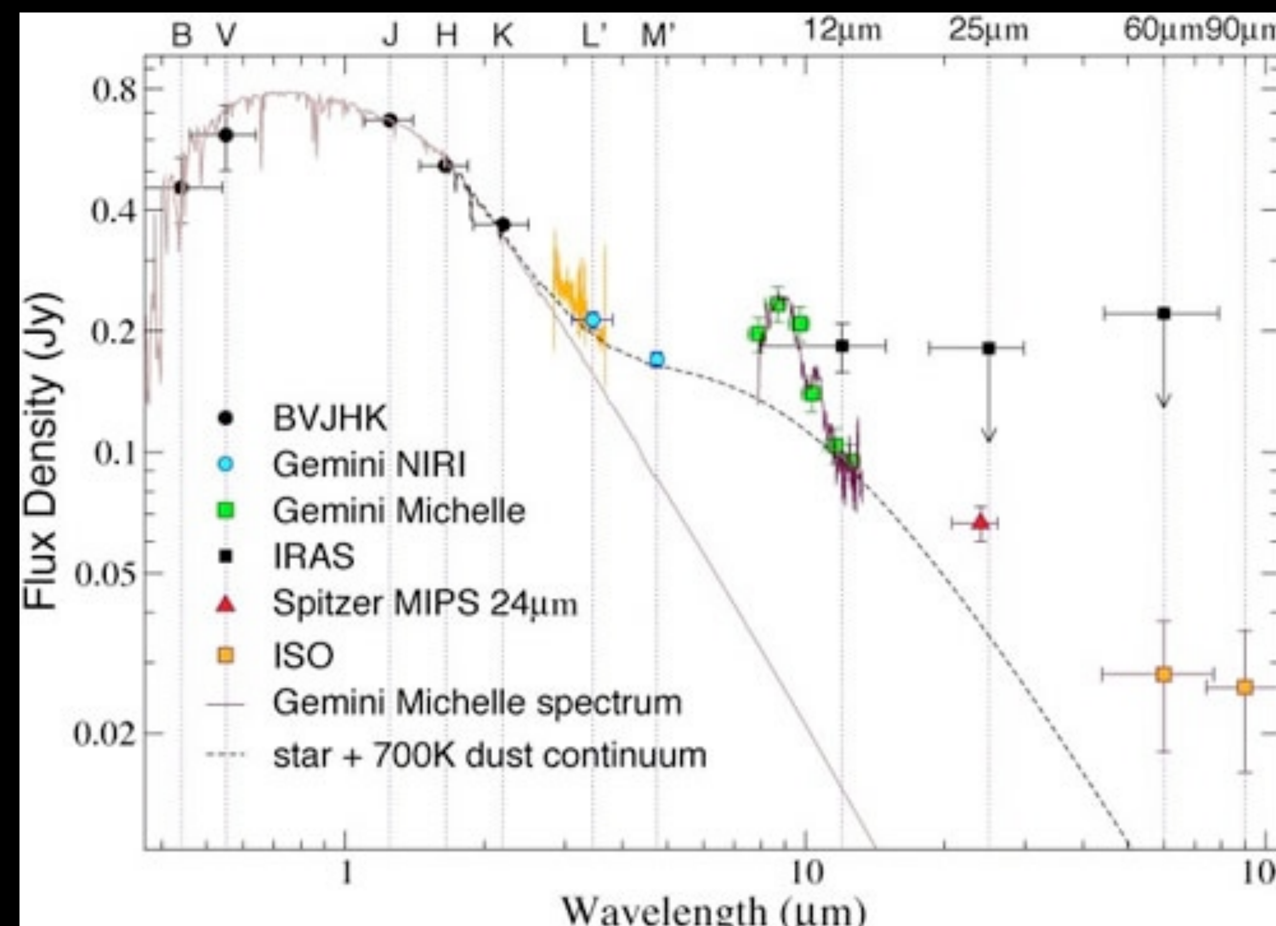
This

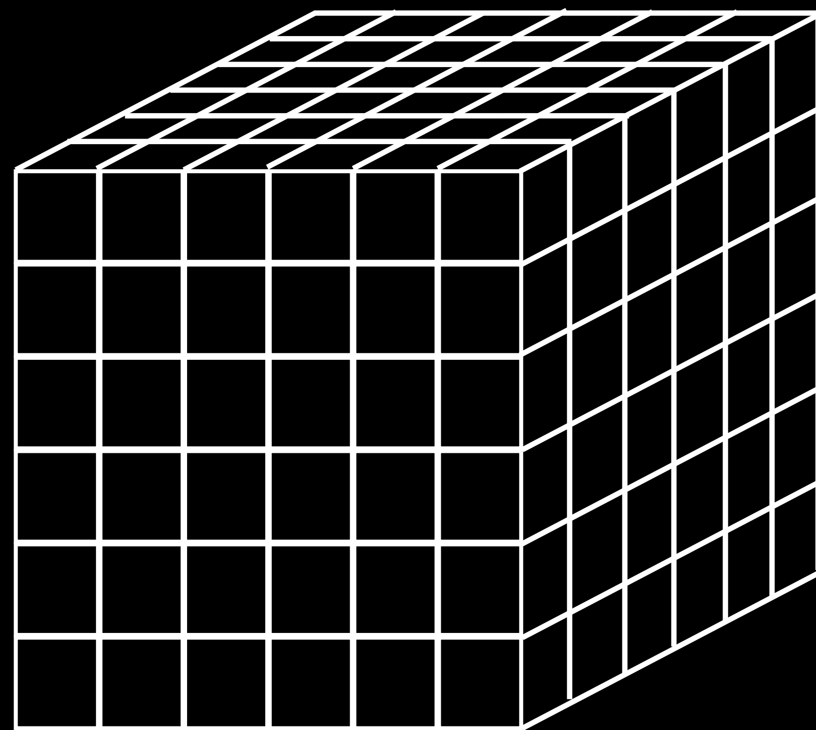

Extraction of
Area of Interest
Weather
Logistics
Coalition Forces
Intelligence, Other Forces
Imagery Overlays
Geospatial Framework

This

is a “spectral energy distribution”







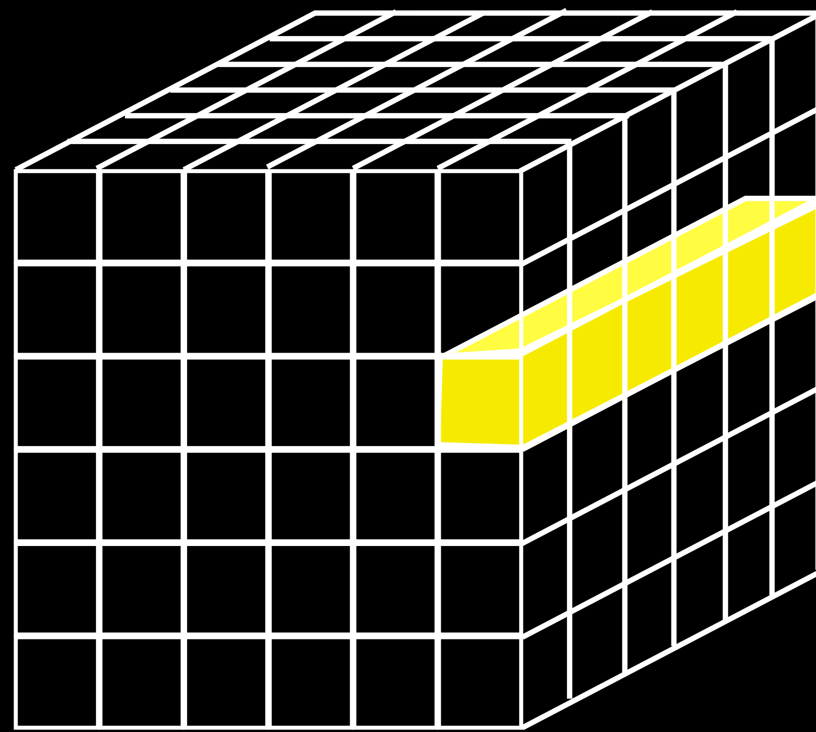
GENERALLY

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"



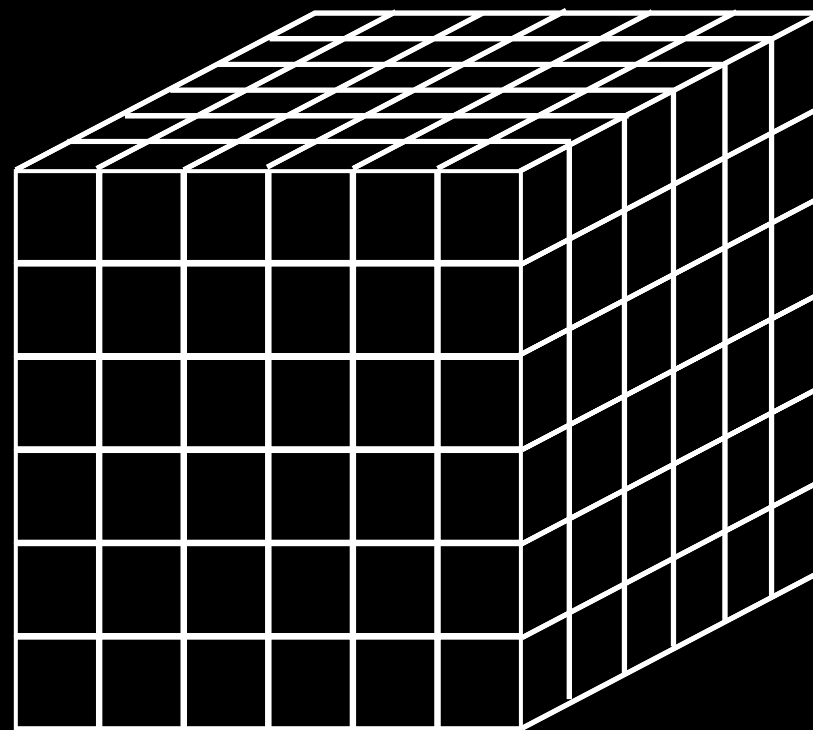
GENERALLY

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”



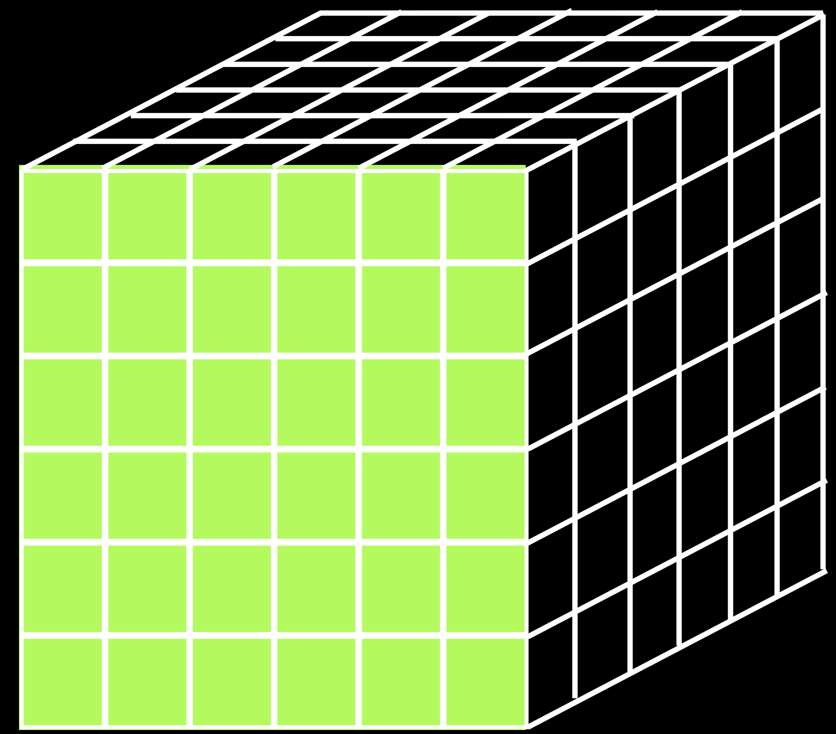
GENERALLY

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”



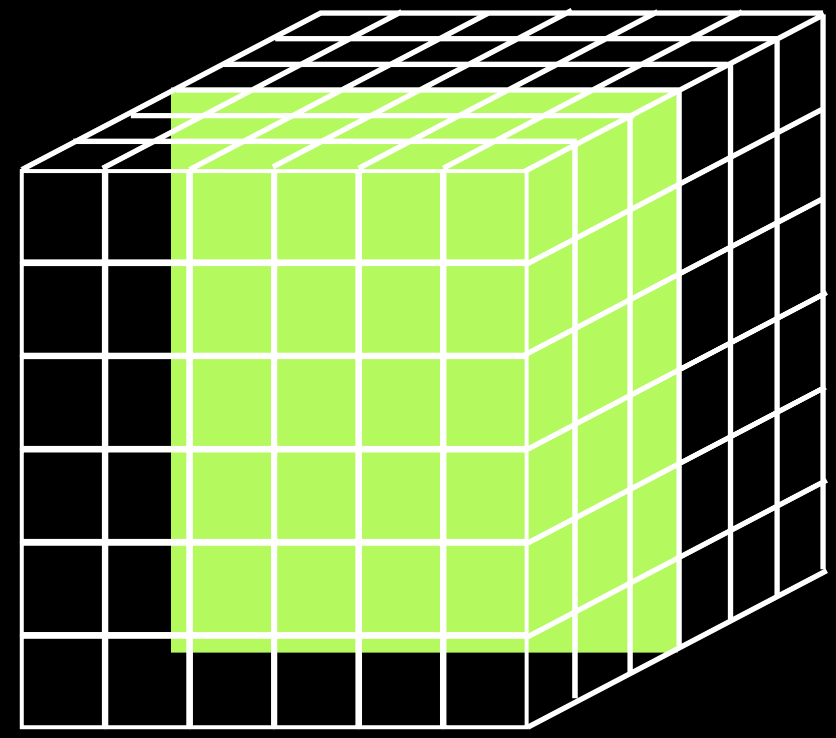
GENERALLY

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”



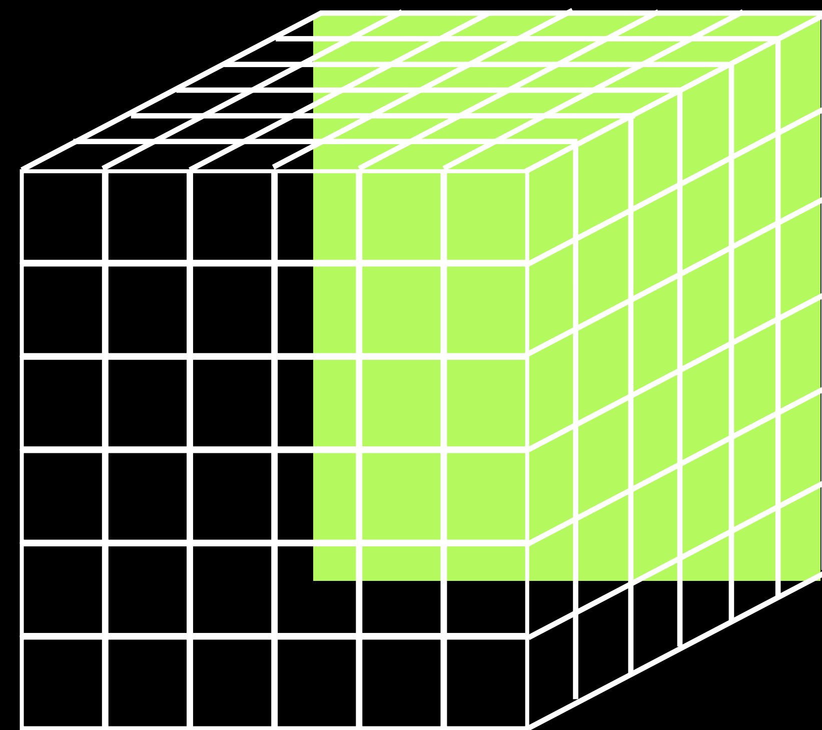
GENERALLY

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”



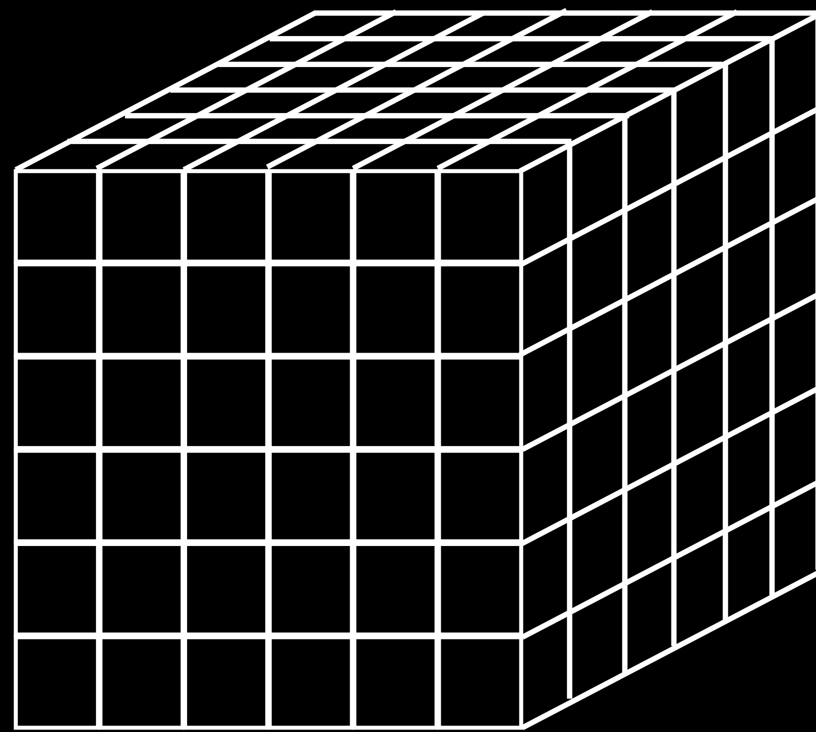
GENERALLY

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”



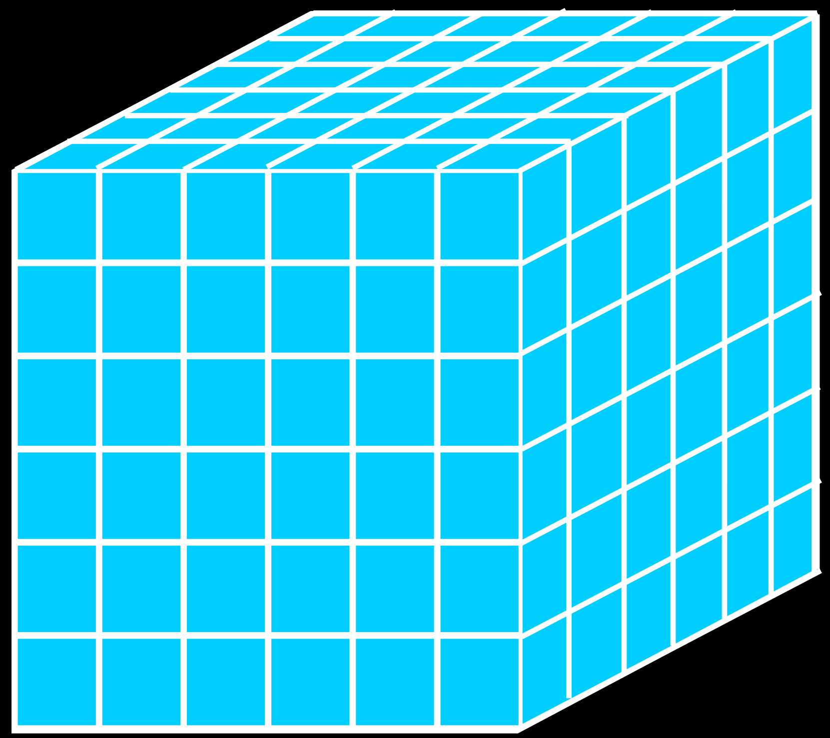
GENERALLY

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”



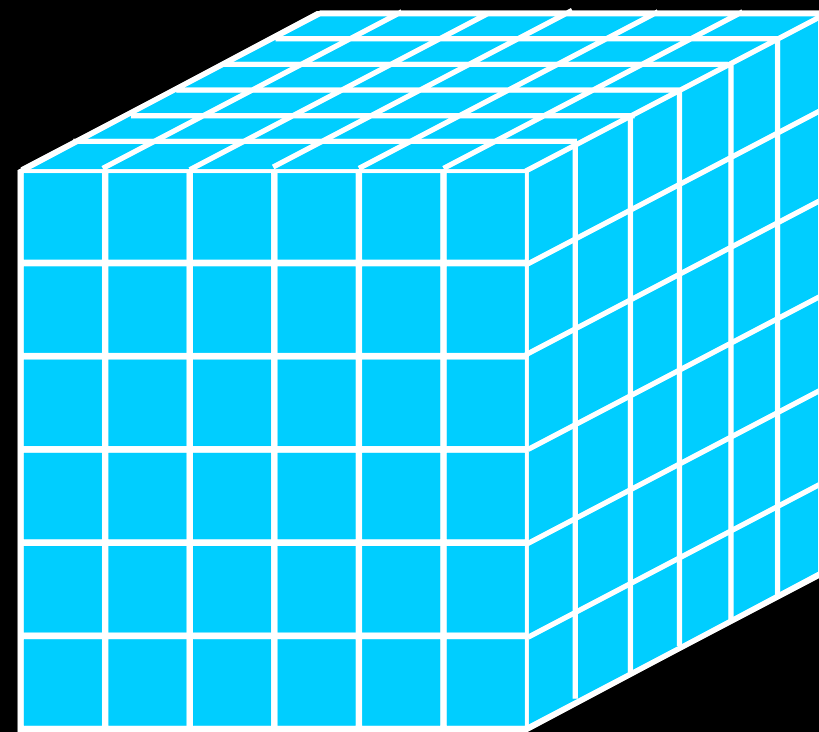
GENERALLY

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”



GENERALLY

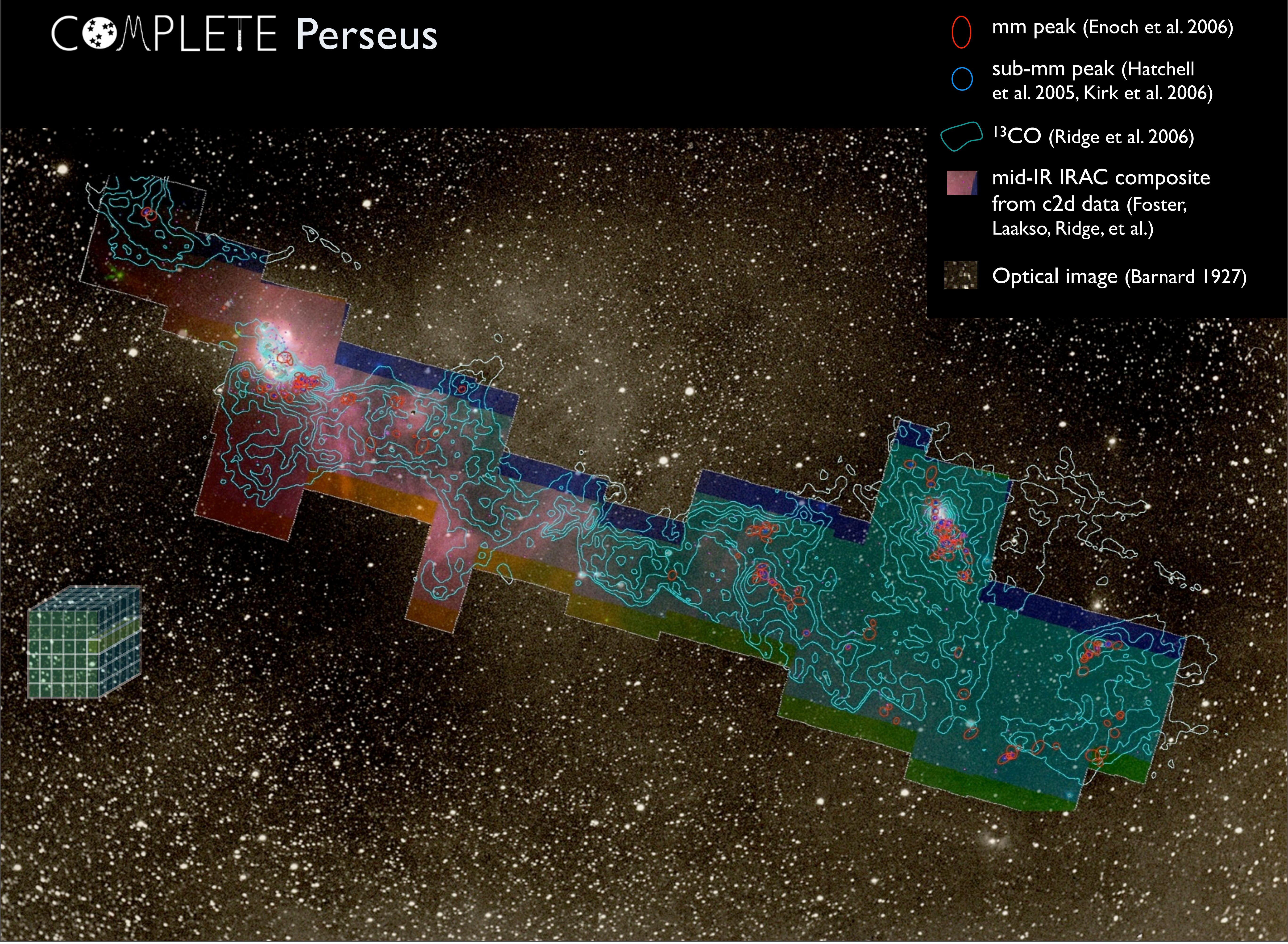
1D: Columns = "Spectra", "SEDs" or "Time Series"

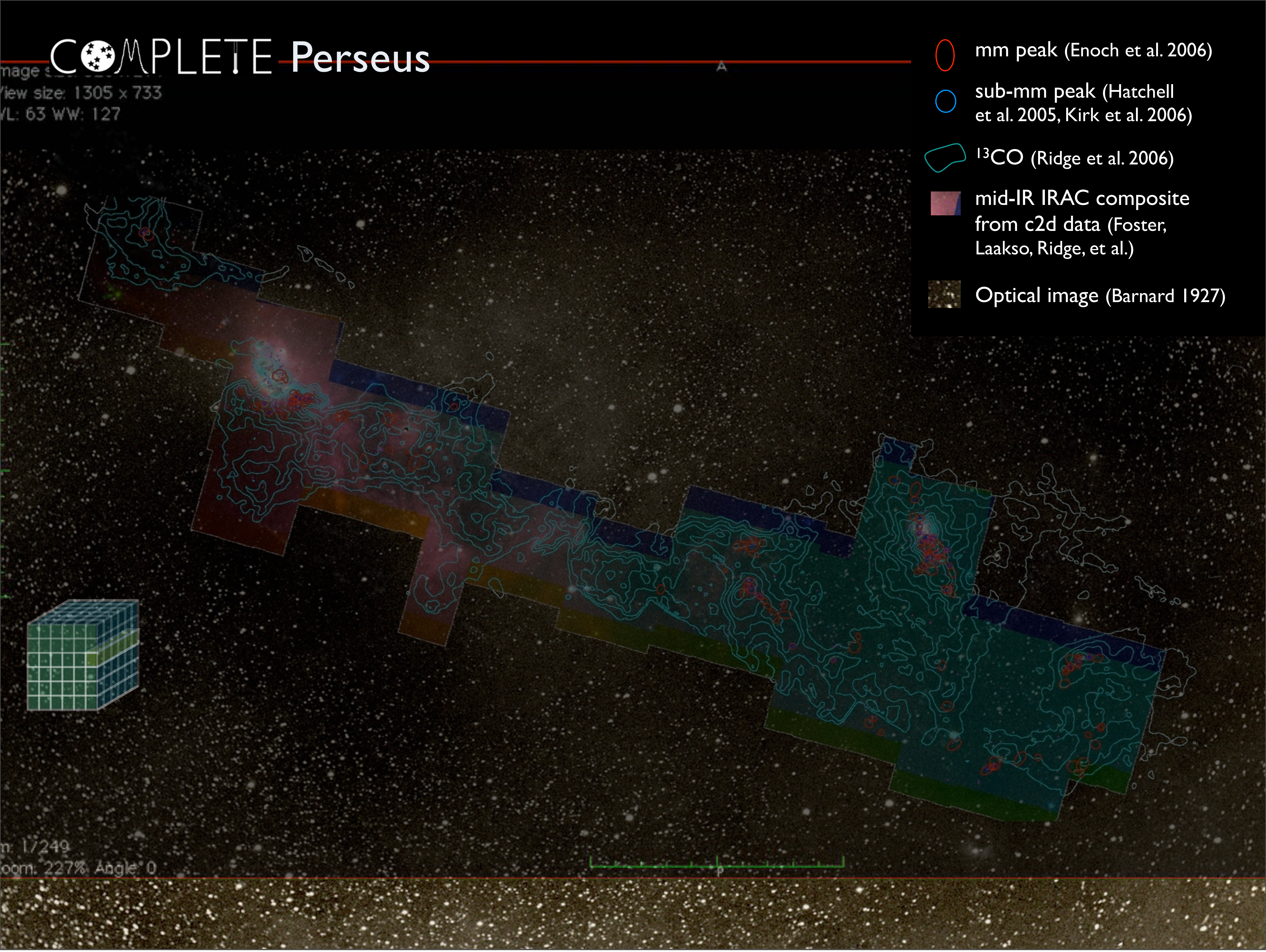
2D: Faces or Slices = "Images"

3D: Volumes = "3D Renderings", "2D Movies"

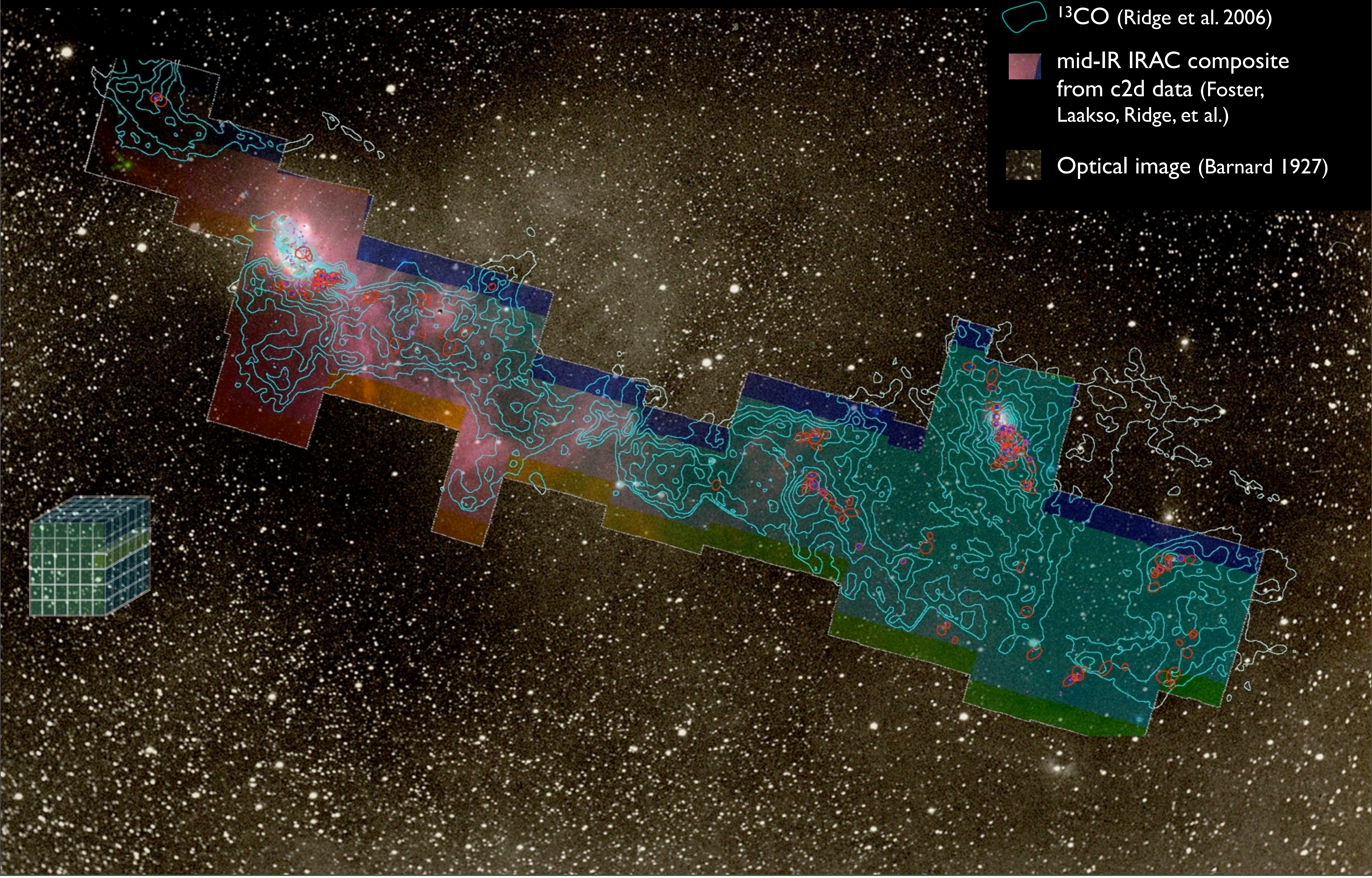
4D: Time Series of Volumes = "3D Movies"

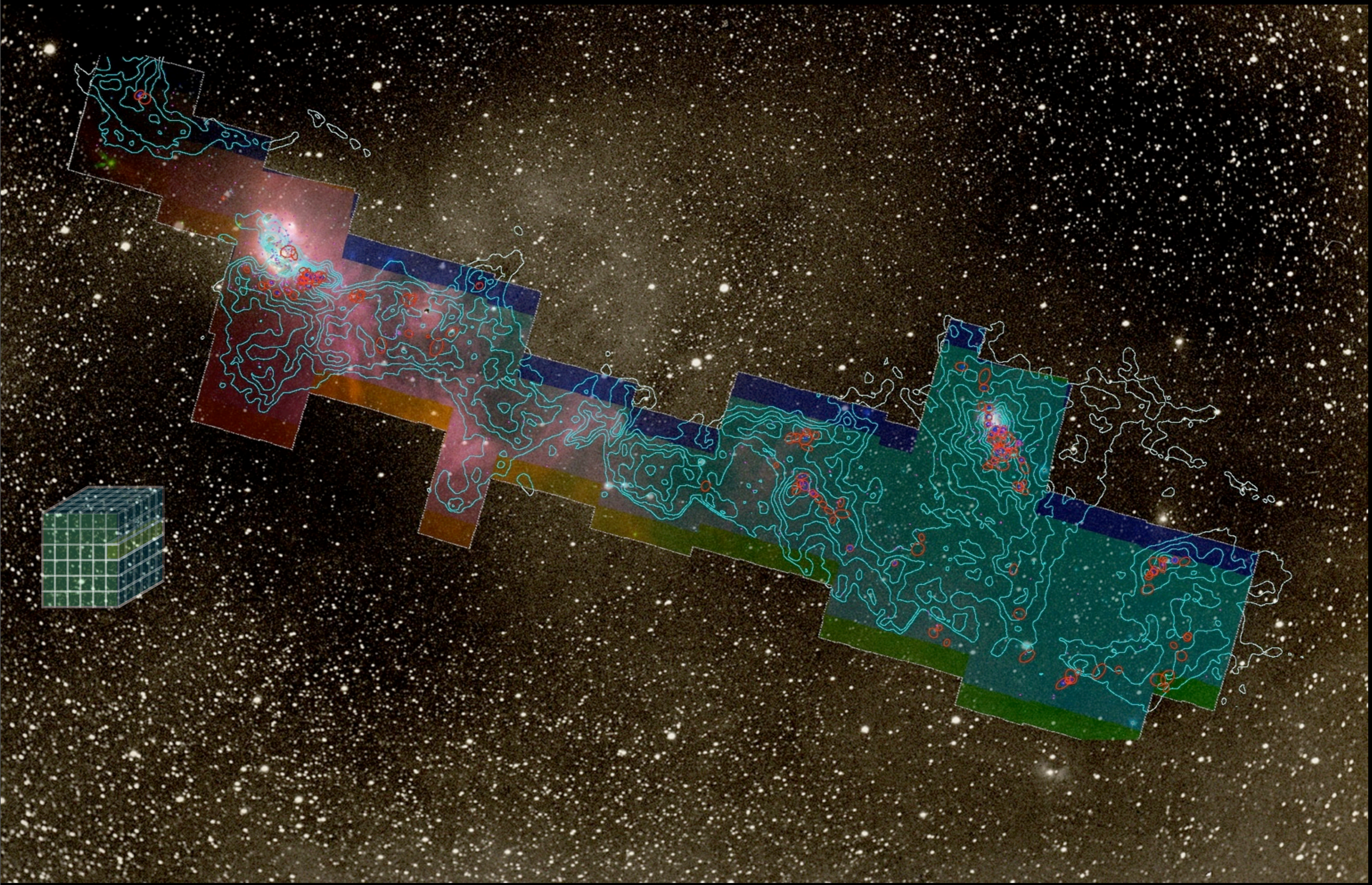
COMPLETE Perseus




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)

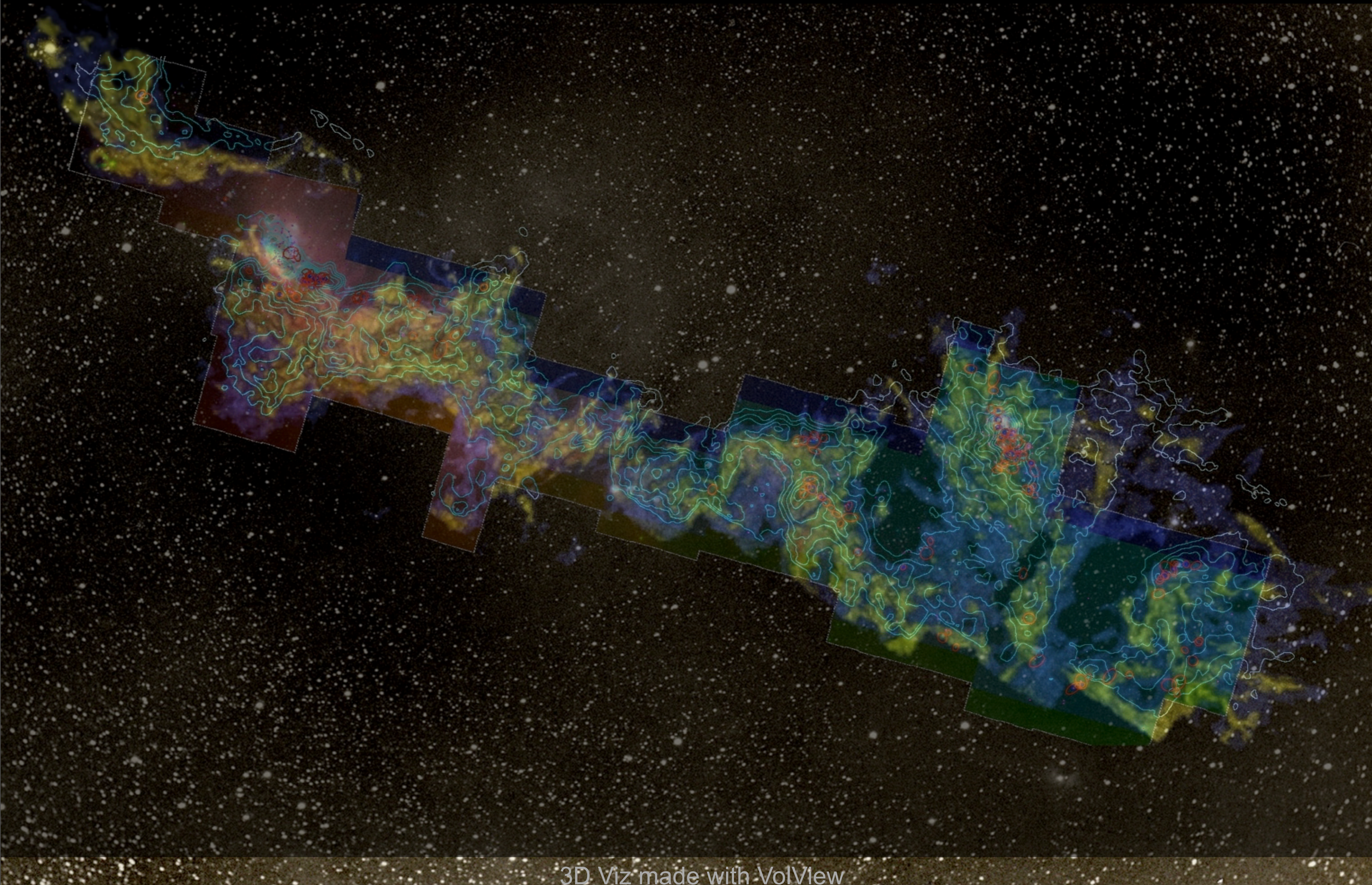
COMPLETE Perseus





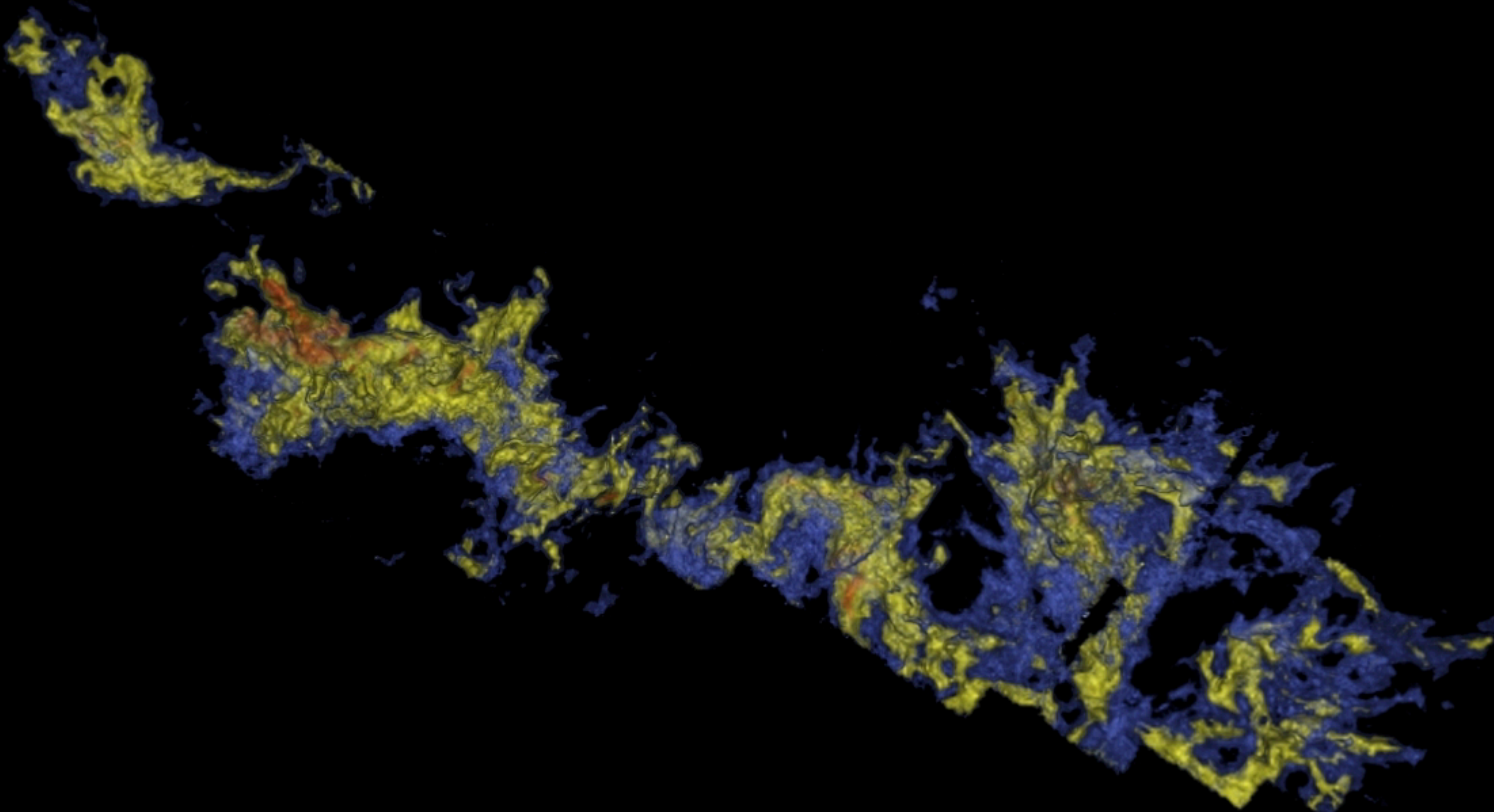
AstronomicalMedicine@IIC

COMPLETE



AstronomicalMedicine@IIC

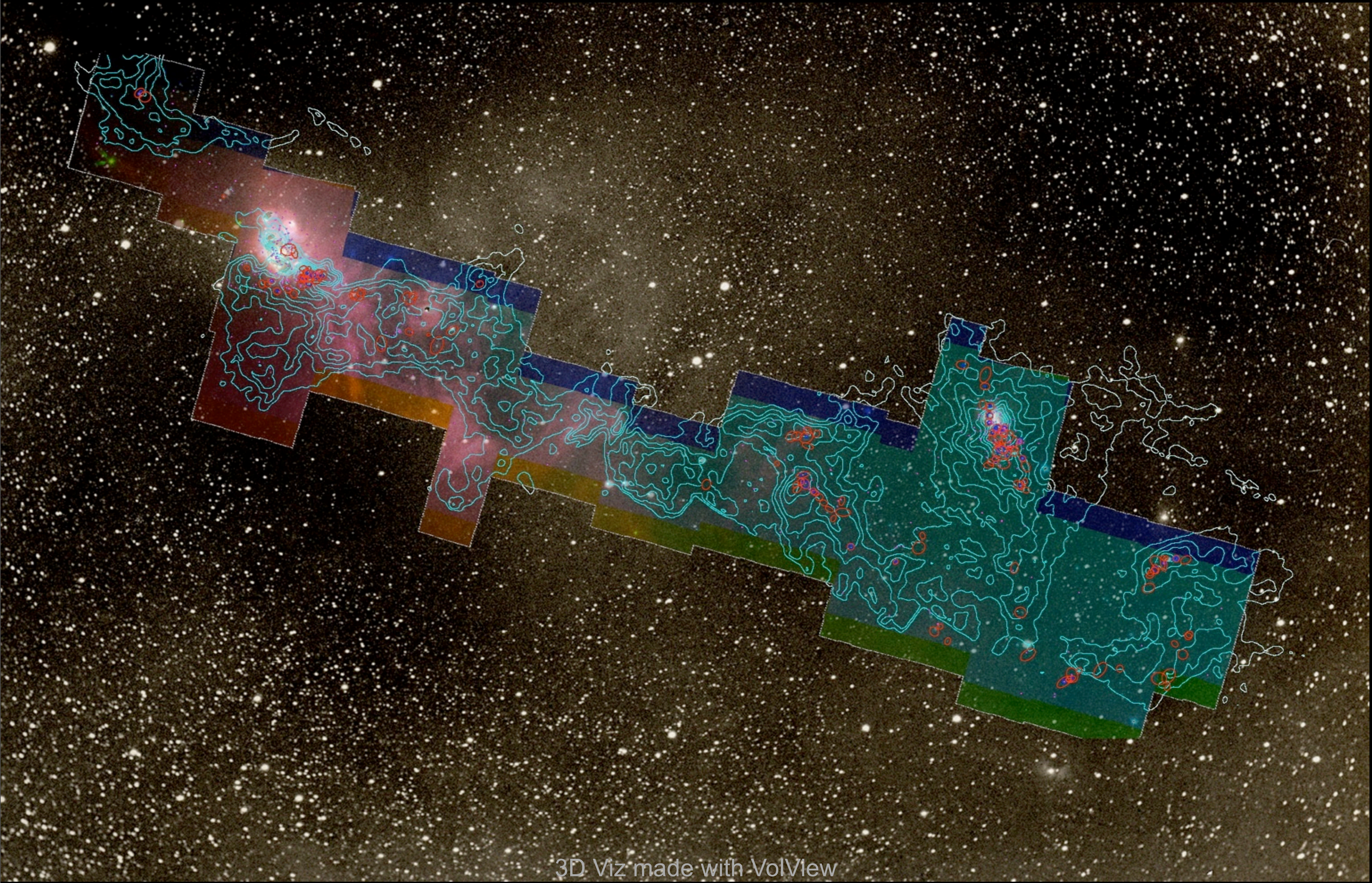
COMPLETE



3D Viz made with VolView

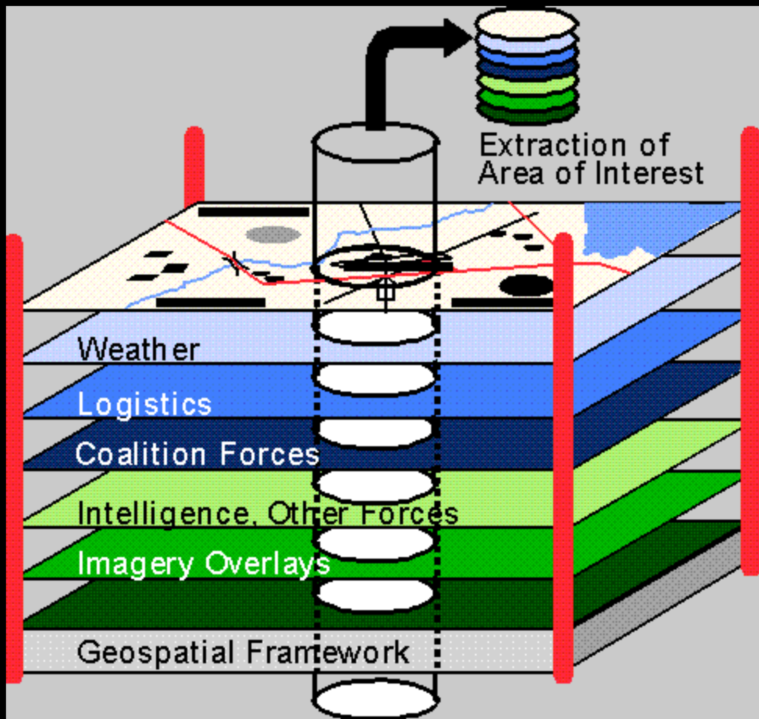
AstronomicalMedicine@IIC

COMPLETE

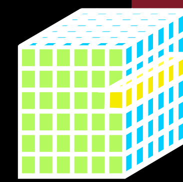


AstronomicalMedicine@IIC

COMPLETE



Data
Dimensions
Display

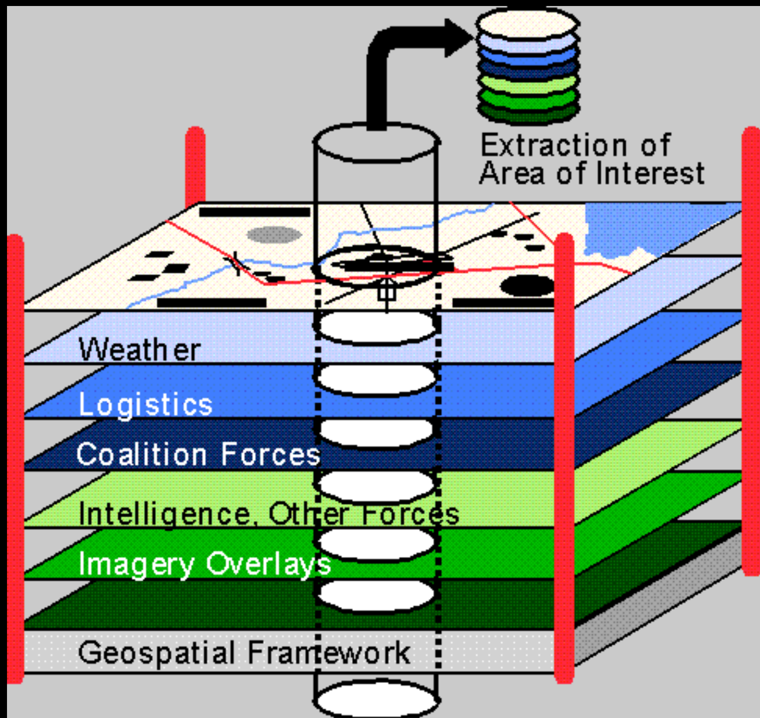


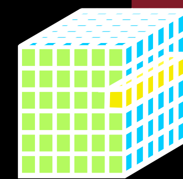
COMPLETE Data Available

Center on Perseus | Center on Ophiuchus | Center on Serpens

Full-Cloud Data (Phase I, All Data Available)

Dataset	Show	Perseus	Ophiuchus	Serpens	Link
GBT: HI Data Cube	☑	√	√	∅	Data
IRAS: Av/Temp Maps	☑	√	√	√	Data
FCRAO: 12CO	☑	√	√	√	Data
FCRAO: 13CO		√	√	√	Data
JCMT: 850 microns	☑	√	√	∅	Data
Spitzer c2d: IRAC 1,3 (3.6,5.8 μm)	☑	√	√	√	Data
Spitzer c2d: IRAC 2,4 (4.5,8 μm)	☑	√	√	√	Data
CSO/Bolocam: 1.2-mm	☑	√	∅	∅	Data
Spitzer MIPS: Derived Dust Map	☑	√	∅	∅	Data
Targeted Regions (Phase II, Some Data Not Yet Available)					
CTIO/Calar Alto: NIR (J,H,Ks)	☑	√	√	∅	Data
IRAM 30-m: N2H+ and C18O	☑	√	∅	∅	Data
IRAM 30-m: 1.1-mm continuum	☑	√	∅	∅	Data
Megacam/MMT: r,i,z images	☑	√	∅	∅	Data
Catalogs & Pointed Surveys					
NH3 Pointed Survey	■	√	∅	∅	Data
YSO Candidate list (c2d)	■	√	√	√	Data

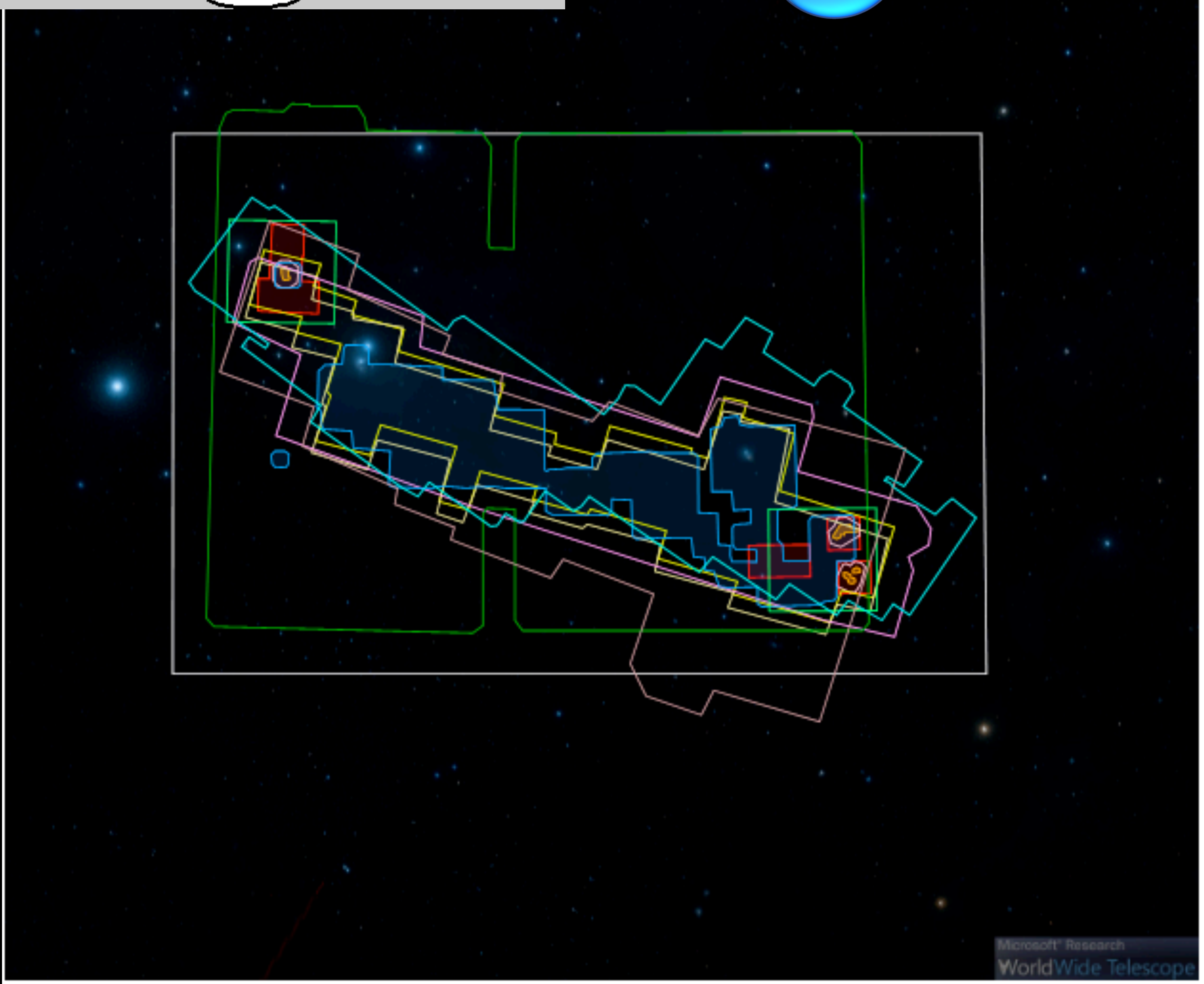




Data Dimensions Display

Open Data Open Tools

COMPLETE

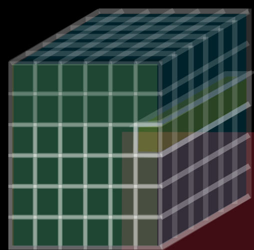


COMPLETE Data Available

Center on Perseus | Center on Ophiuchus | Center on Serpens

Full-Cloud Data (Phase I, All Data Available)

Dataset	Show	Perseus	Ophiuchus	Serpens	Link
GBT: HI Data Cube		√	√	∅	Data
IRAS: Av/Temp Maps		√	√	√	Data
FCRAO: 12CO		√	√	√	Data
FCRAO: 13CO		√	√	√	Data
JCMT: 850 microns		√	√	∅	Data
Spitzer c2d: IRAC 1,3 (3.6,5.8 μm)		√	√	√	Data
Spitzer c2d: IRAC 2,4 (4.5,8 μm)		√	√	√	Data
CSO/Bolocam: 1.2-mm		√	∅	∅	Data
Spitzer MIPS: Derived Dust Map		√	∅	∅	Data
Targeted Regions (Phase II, Some Data Not Yet Available)					
CTIO/Calar Alto: NIR (J,H,Ks)		√	√	∅	Data
IRAM 30-m: N2H+ and C18O		√	∅	∅	Data
IRAM 30-m: 1.1-mm continuum		√	∅	∅	Data
Megacam/MMT: r,i,z images		√	∅	∅	Data
Catalogs & Pointed Surveys					
NH3 Pointed Survey		√	∅	∅	Data
YSO Candidate list (c2d)		√	√	√	Data



Data
Dimensions
Display

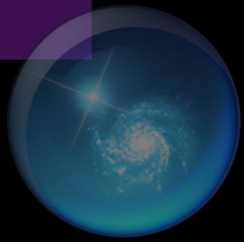
Linked Views

"Taste-Testing"
[Sim:Data]

Seamless
Astronomy
[the future]

ADS Labs
[info viz]

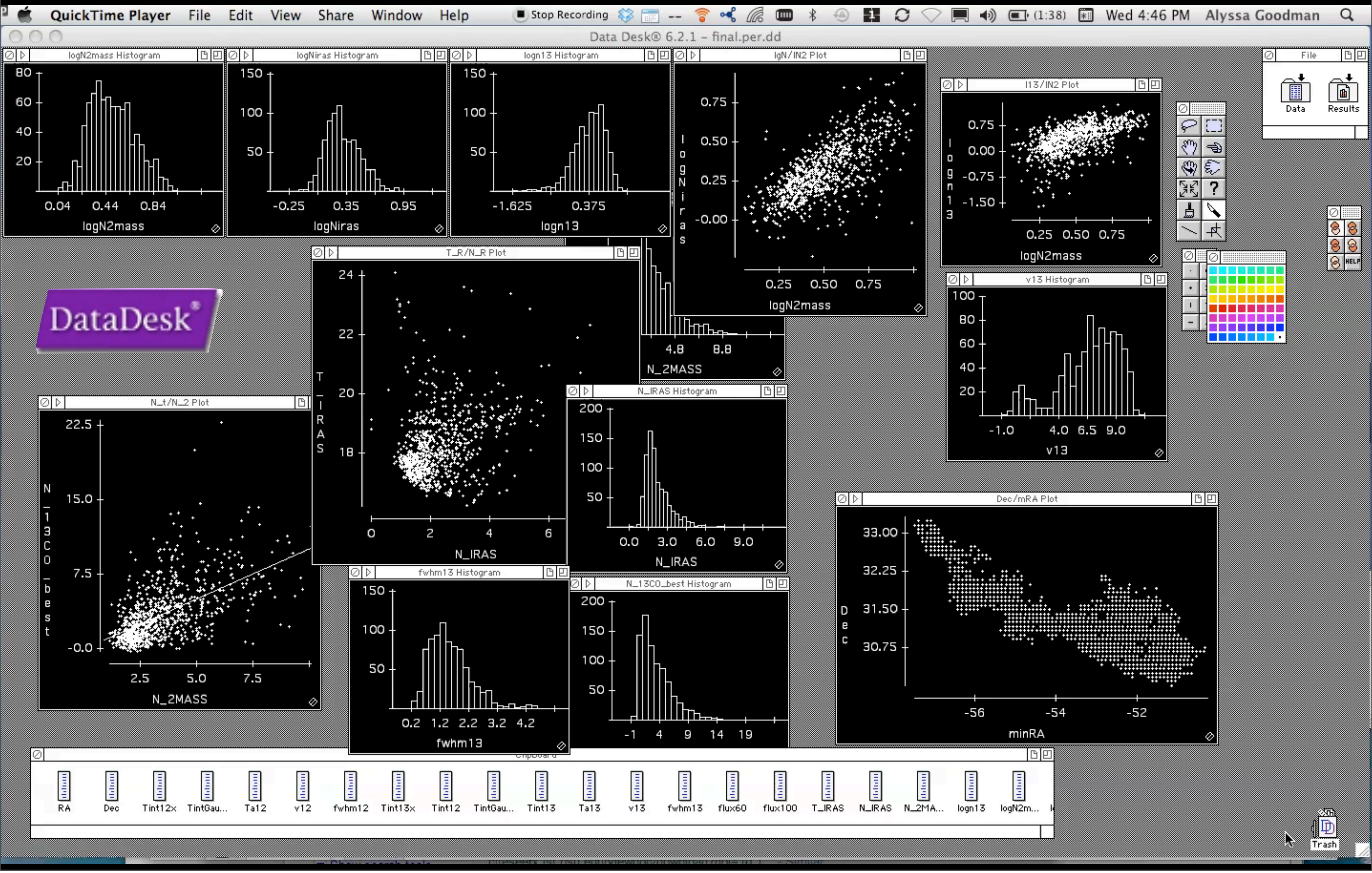
Open Data
Open Tools



DataDesk (est. 1986)

DataDesk®

DataDesk (est. 1986)



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

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

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

Picturing

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

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

Picturing

Rotation

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

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

Picturing

Rotation

Isolation

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

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

Picturing

Rotation

Isolation

Masking

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

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

Picturing

Rotation

Isolation

Masking

Selection

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

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

Picturing

Rotation

Isolation

Masking

Selection

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

Brushing

Linking

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

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)

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

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

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

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

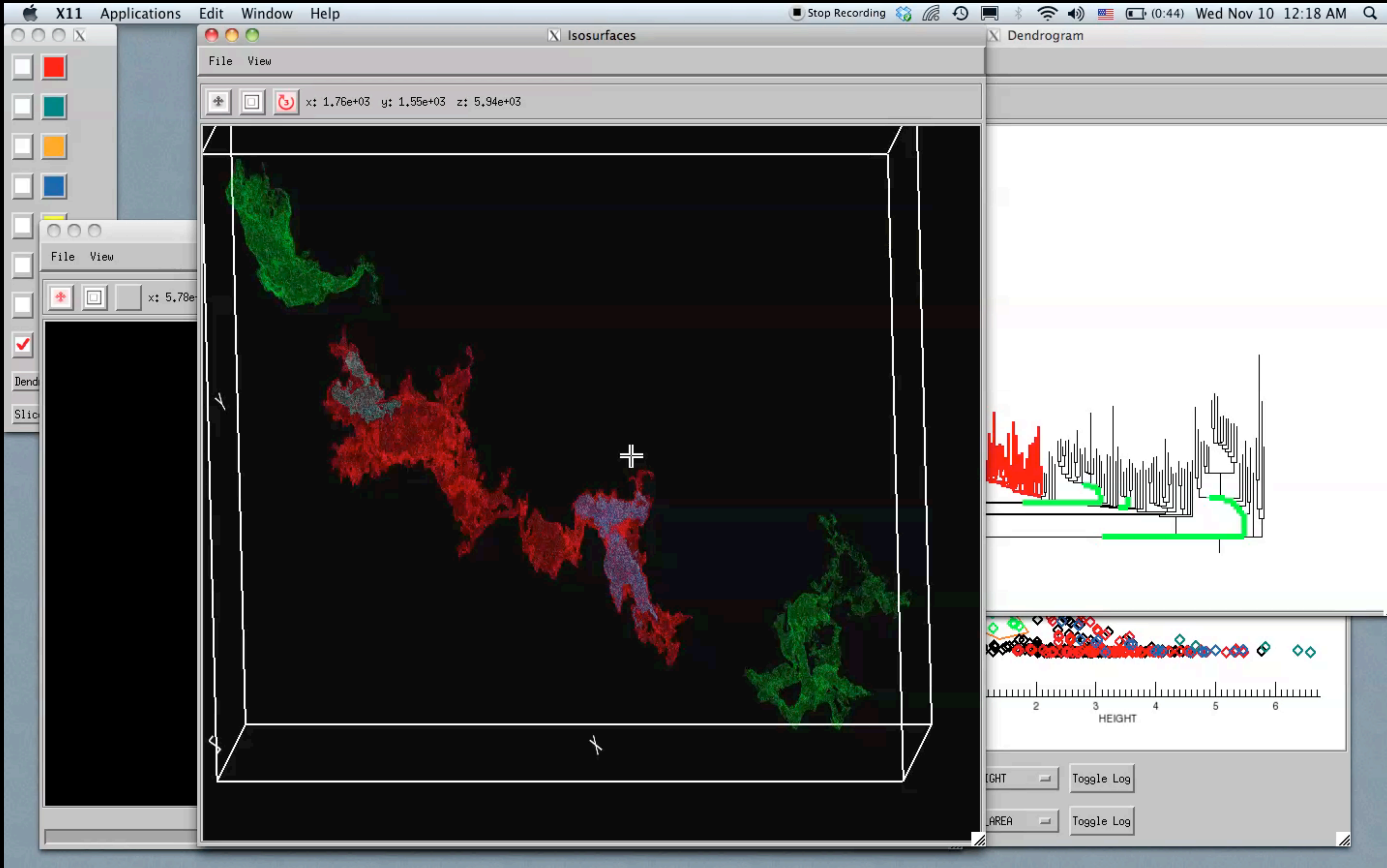
Key words cosmology: large-scale structure – ISM: clouds – methods: data analysis – techniques: image processing – techniques: radial velocities

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.

Exemplar: Linked Dendrogram Views in IDL

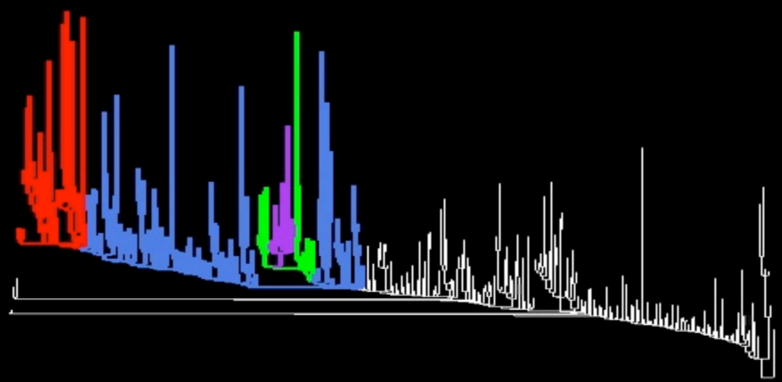
Video & implementation: Christopher Beaumont, CfA/UHawaii;
inspired by AstroMed work of Douglas Alan, Michelle Borkin, AG, Michael Halle, Erik Rosolowsky

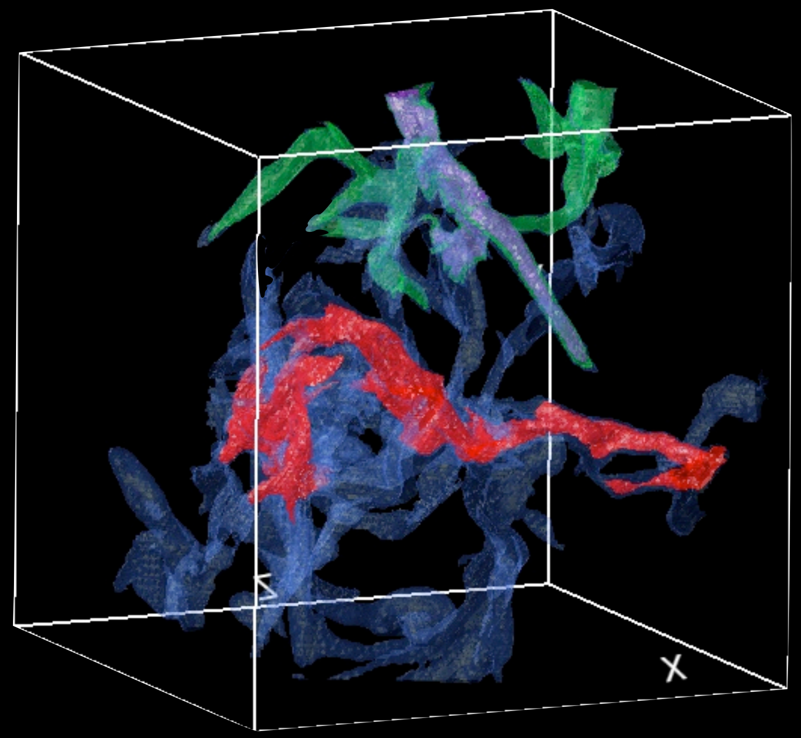
Exemplar: Linked Dendrogram Views in IDL



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

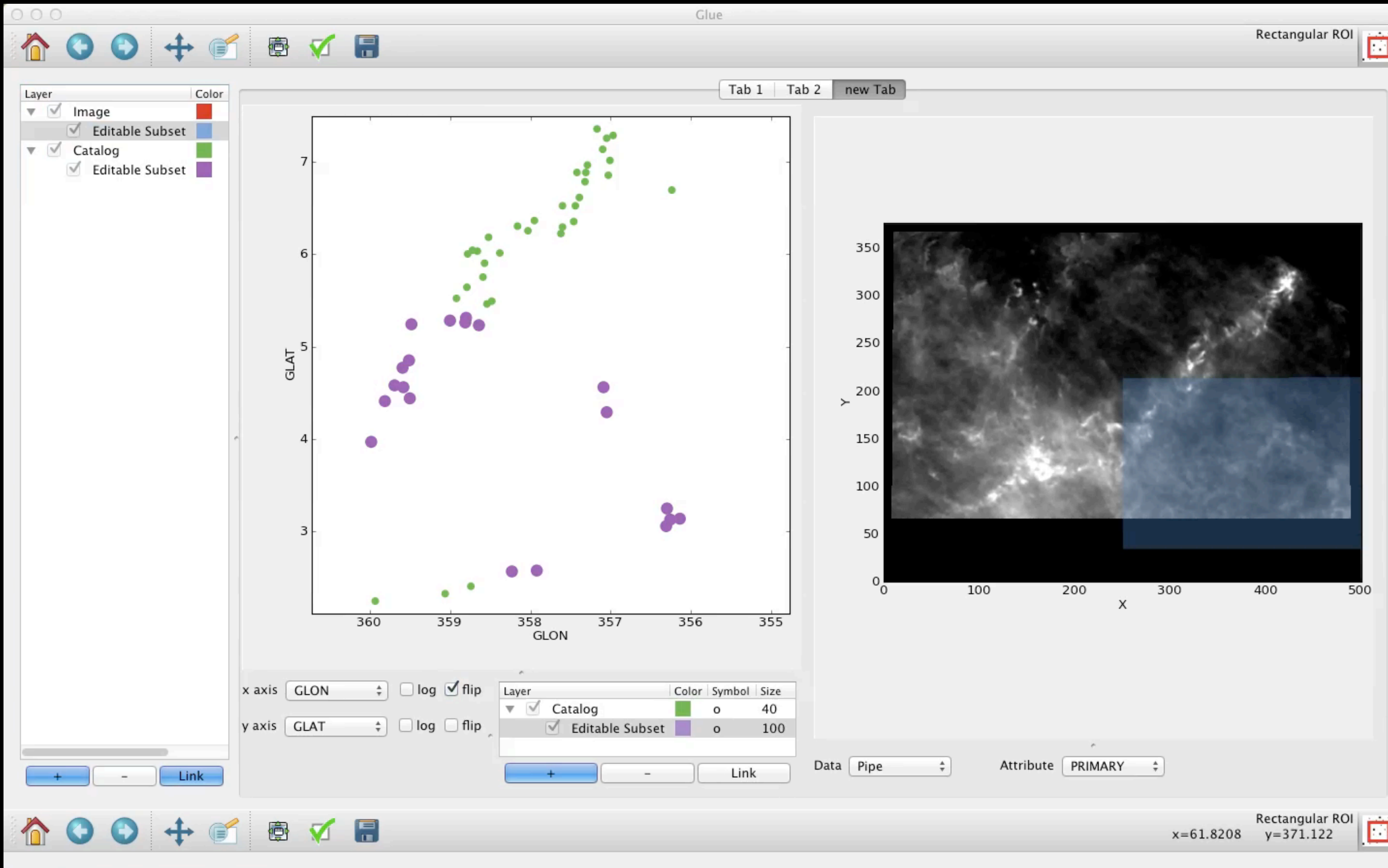
Glue





Current linked view work by Beaumont, Borkin, Goodman, Pfister & Robitaille

Glue



Current linked view work by Beaumont, Borkin, Goodman, Pfister & Robitaille

Value of "Linked Views" from Harvard Undergrad Thesis* Work

on "A Hierarchical Catalog of Molecular Clouds in the Milky Way"

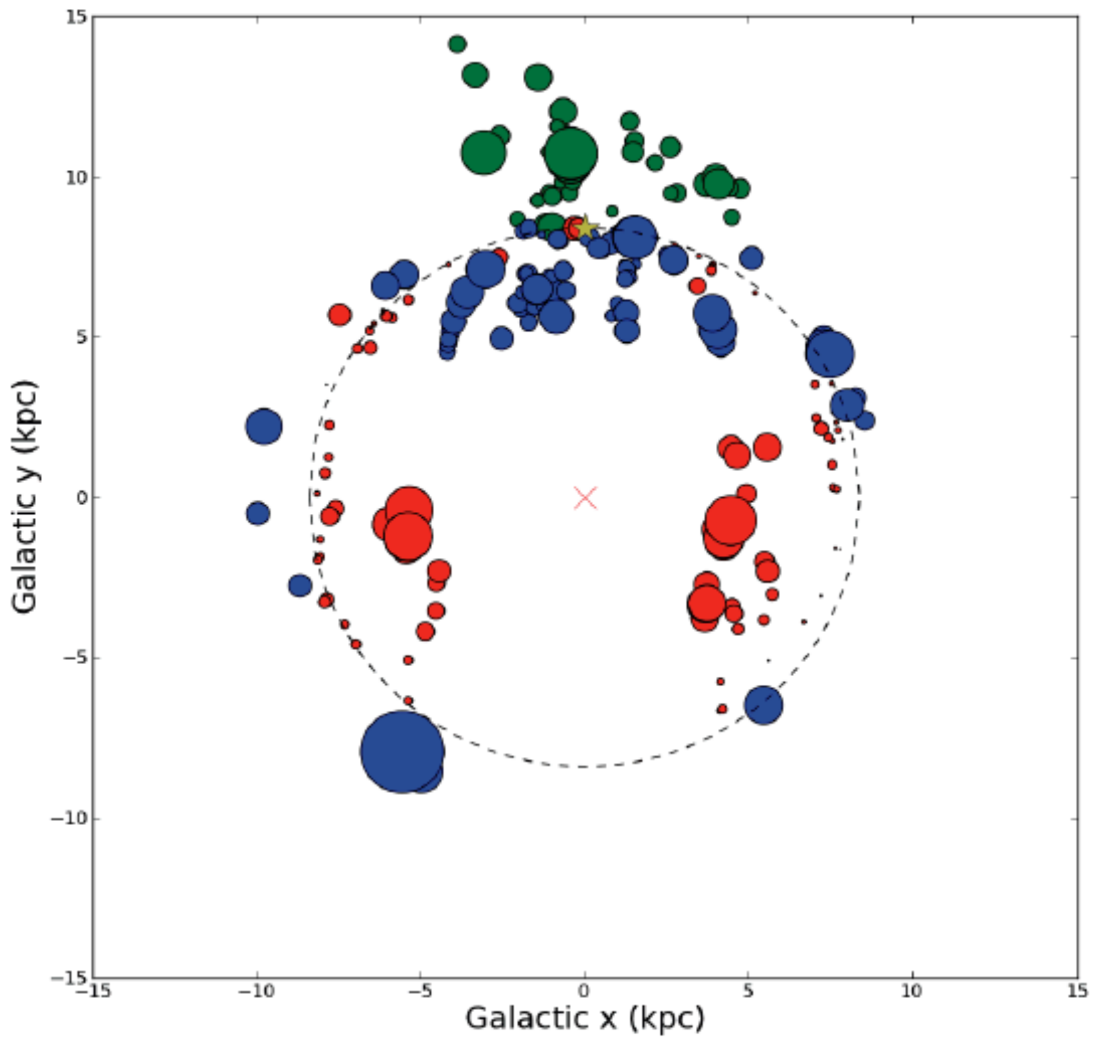
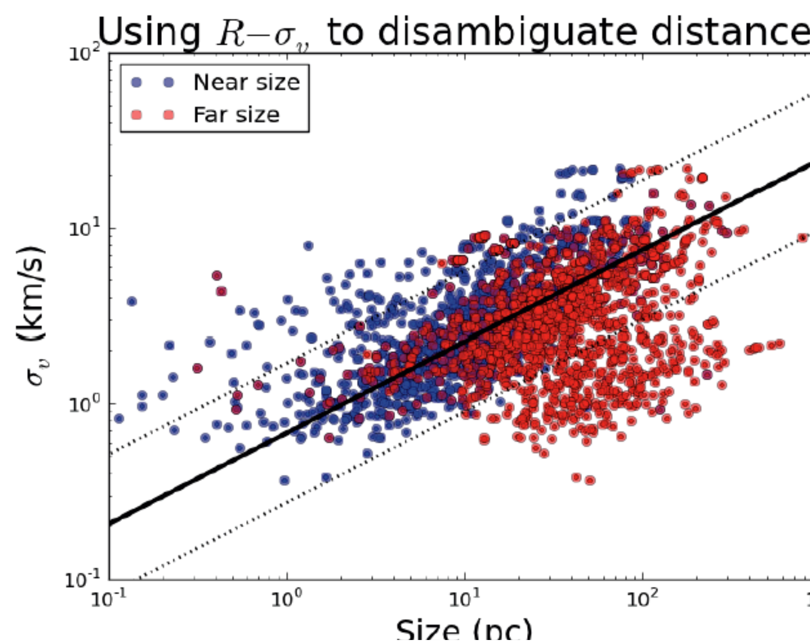


Fig. 8.— All of the "bound clouds" identified in our catalog, plotted face-on in Galactocentric X, Y coordinates. The Sun is located at the yellow star in the upper center. Green: Outer Galaxy; Blue: Near Distance or single-distance solution; Red: Far distance.



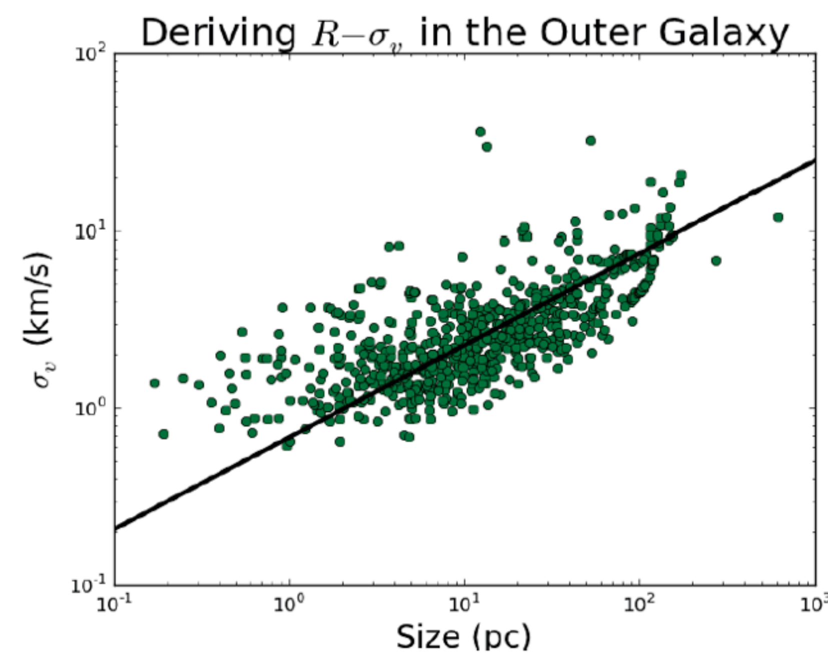
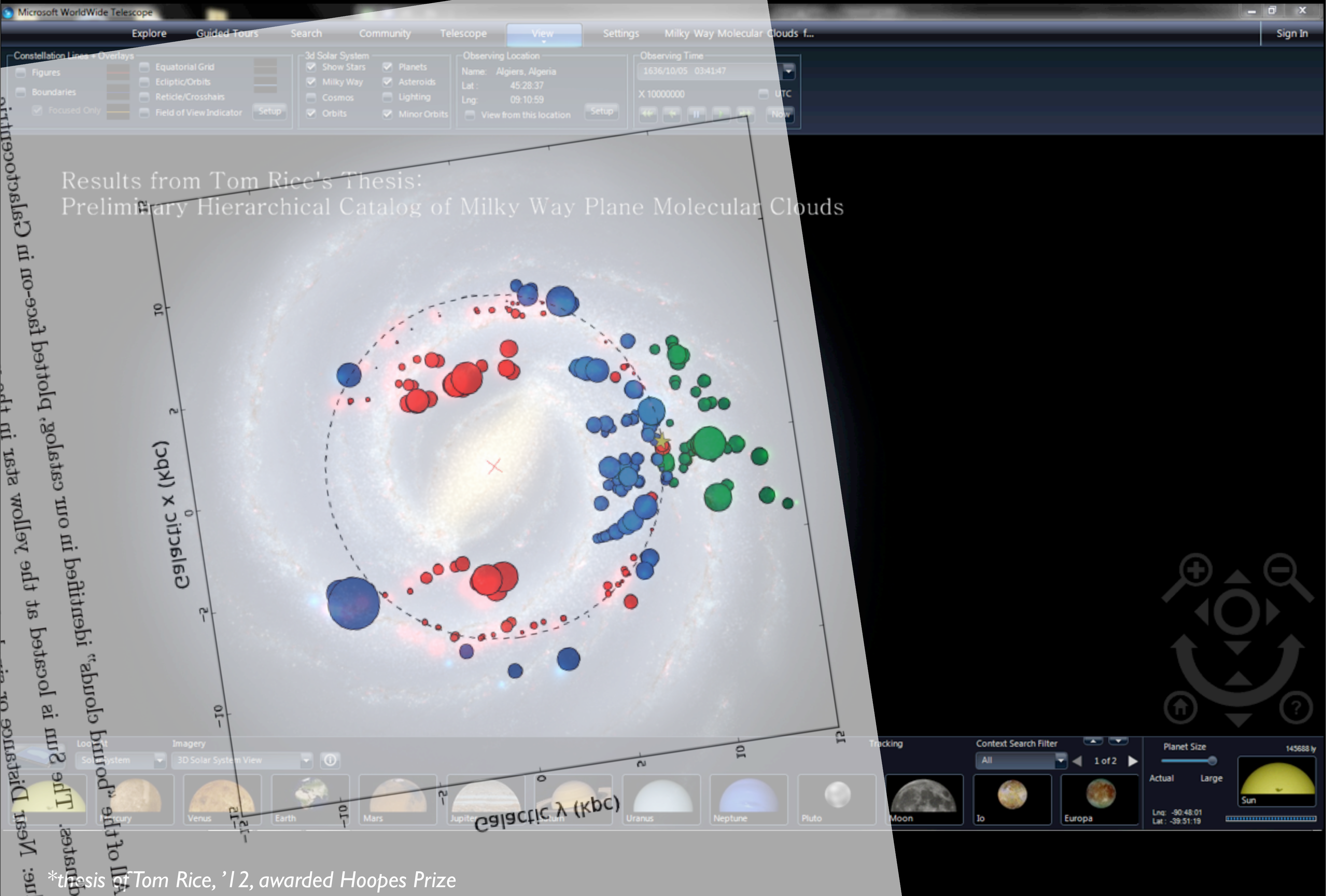


Fig. 5.— The size-linewidth relationship for clouds in the Outer Galaxy. We obtained a fit of $\sigma_v = 0.68\sigma_R^0.52$.

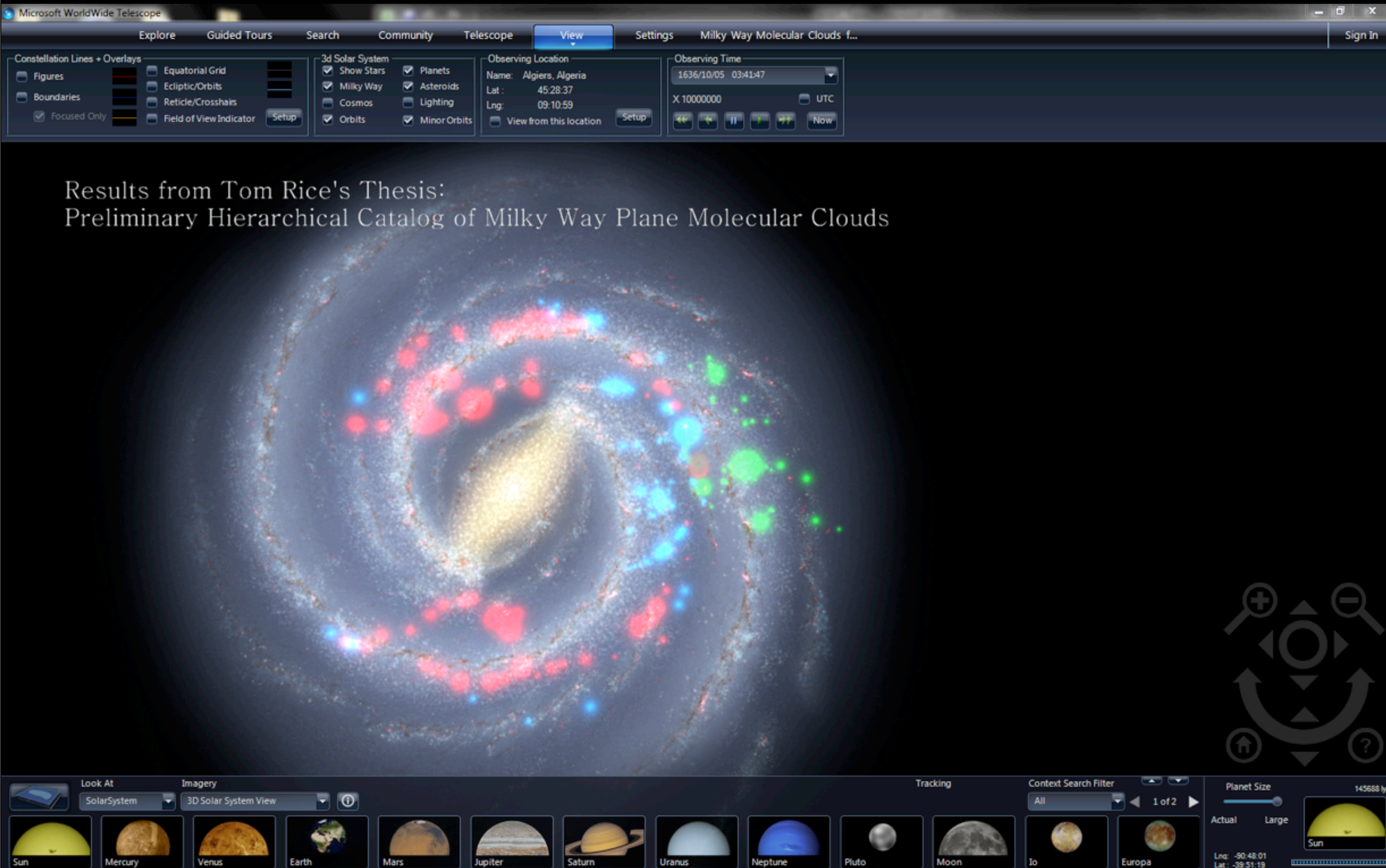
**thesis of Tom Rice, '12, awarded Hoopes Prize*

Value of "Linked Views" from Harvard Undergrad Thesis* Work



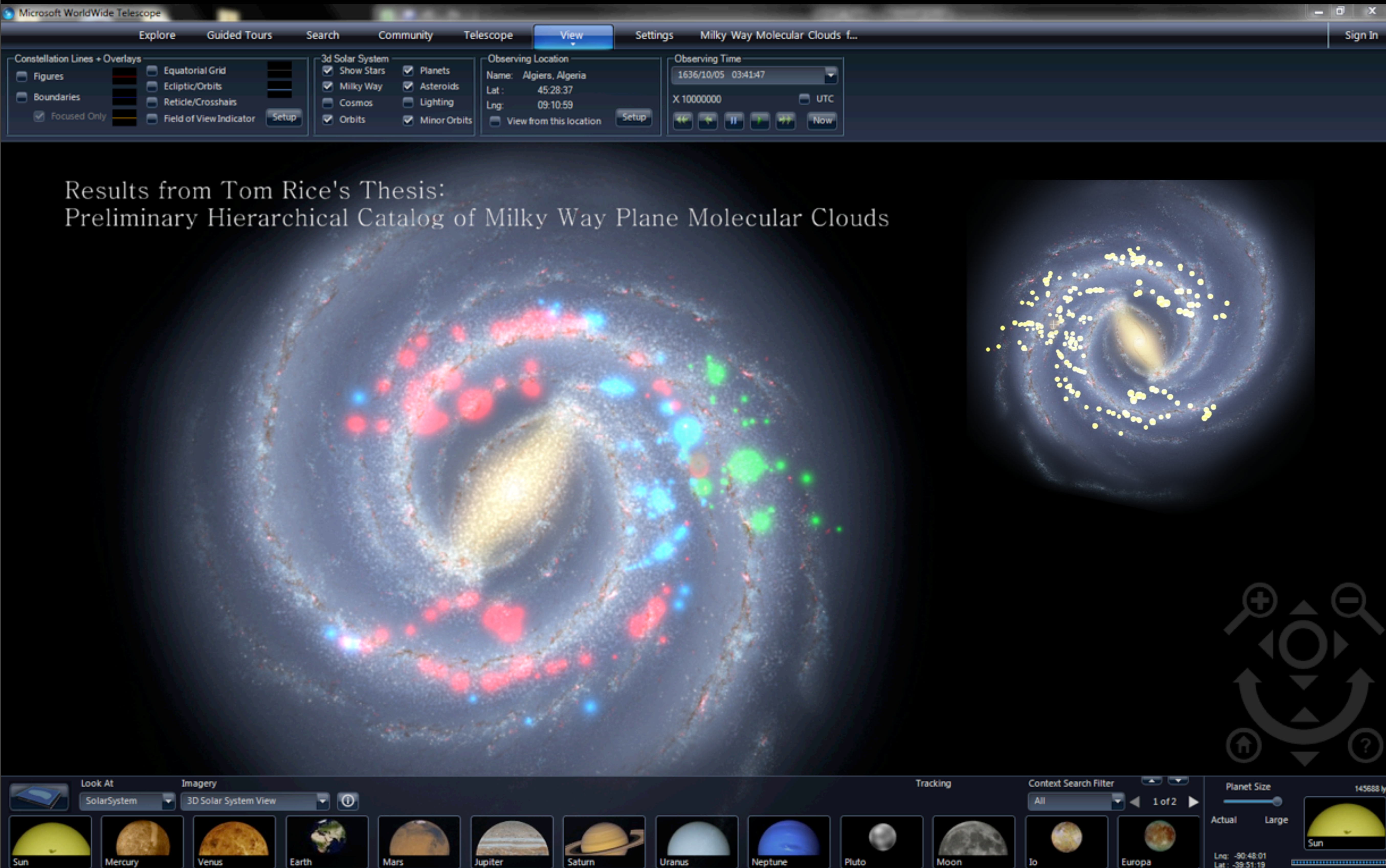
*thesis of Tom Rice, '12, awarded Hoopes Prize

Value of “Linked Views” from Harvard Undergrad Thesis* Work



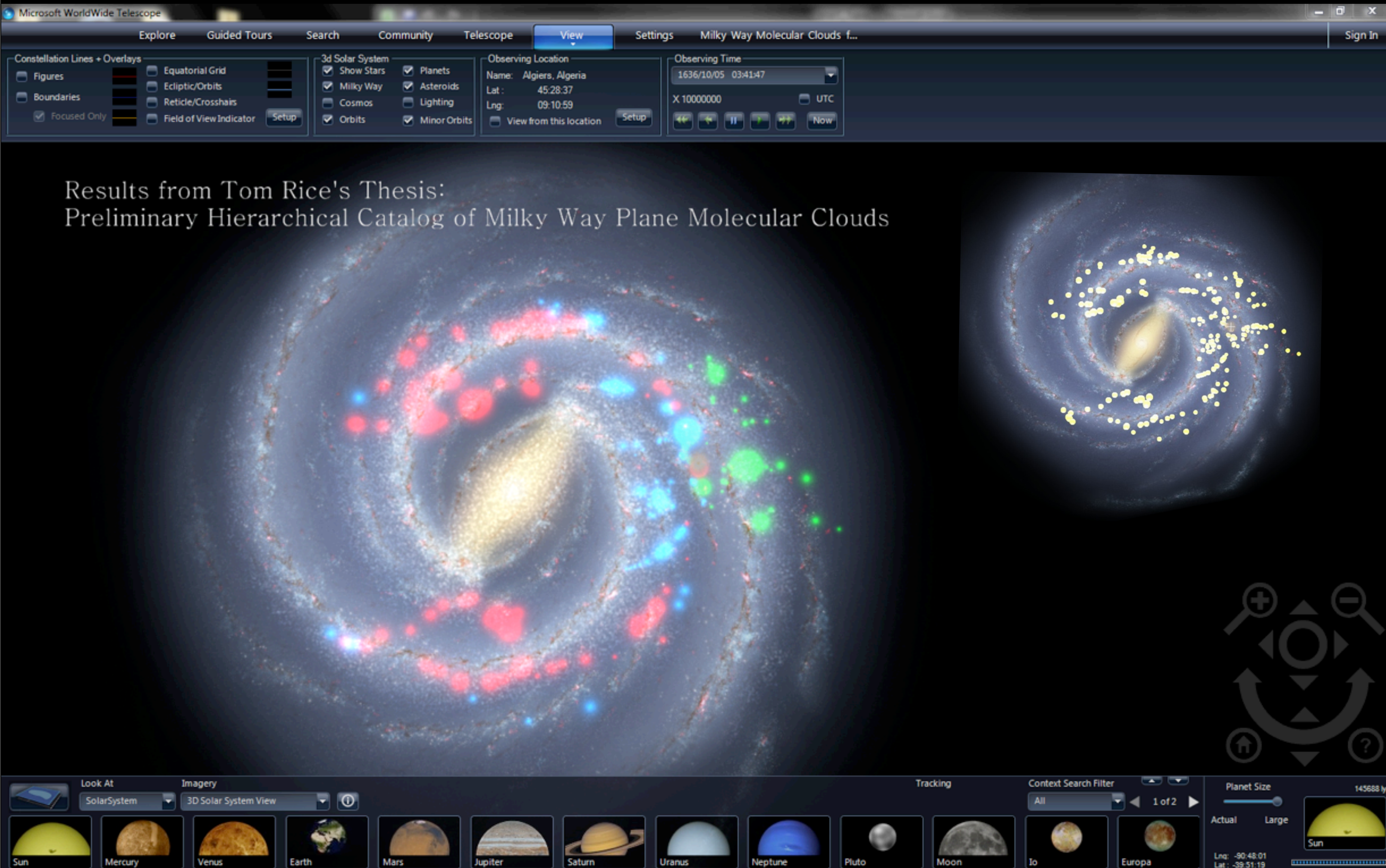
**thesis of Tom Rice, ’12, awarded Hoopes Prize*

Value of “Linked Views” from Harvard Undergrad Thesis* Work



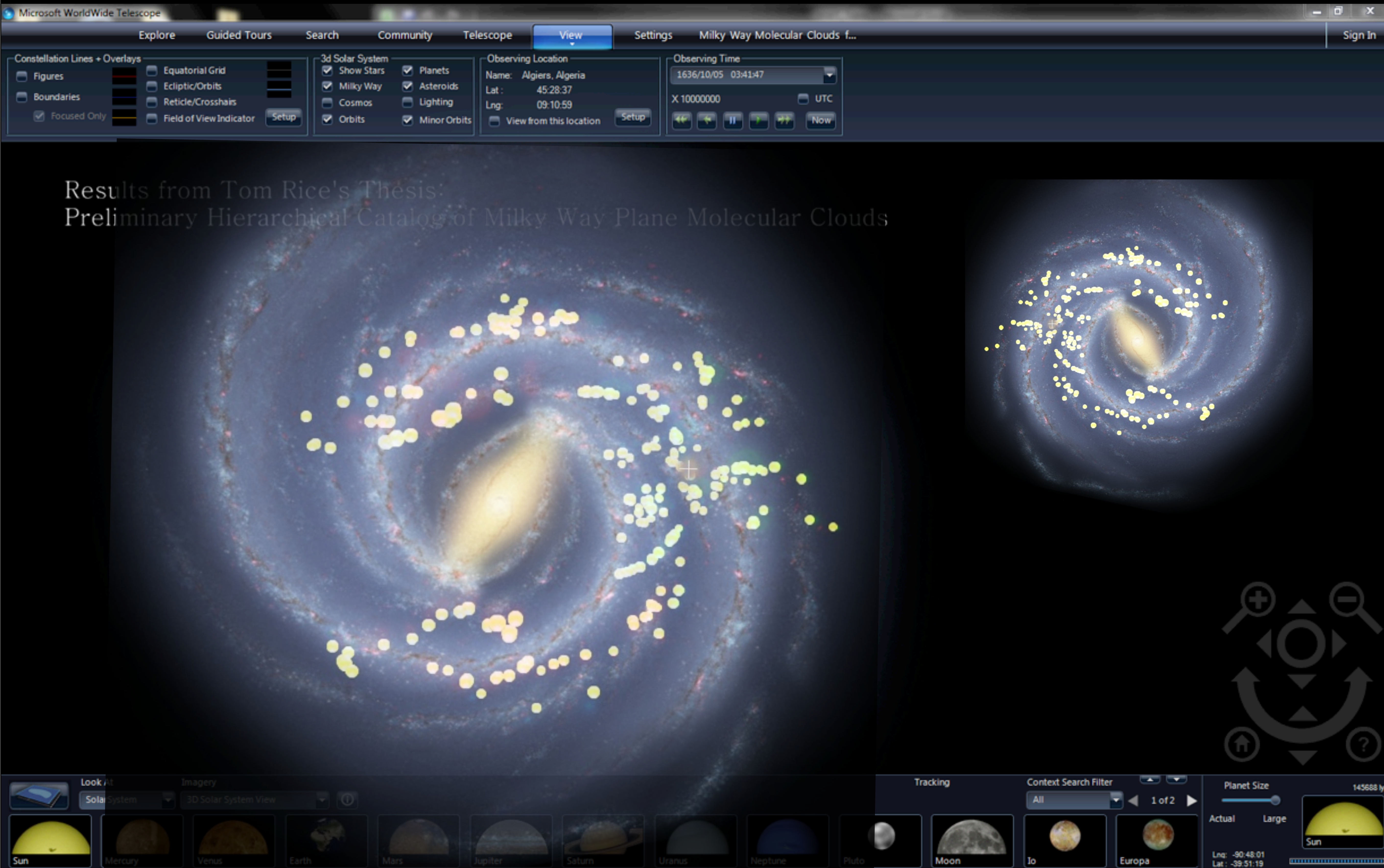
*thesis of Tom Rice, ’12, awarded Hoopes Prize

Value of "Linked Views" from Harvard Undergrad Thesis* Work

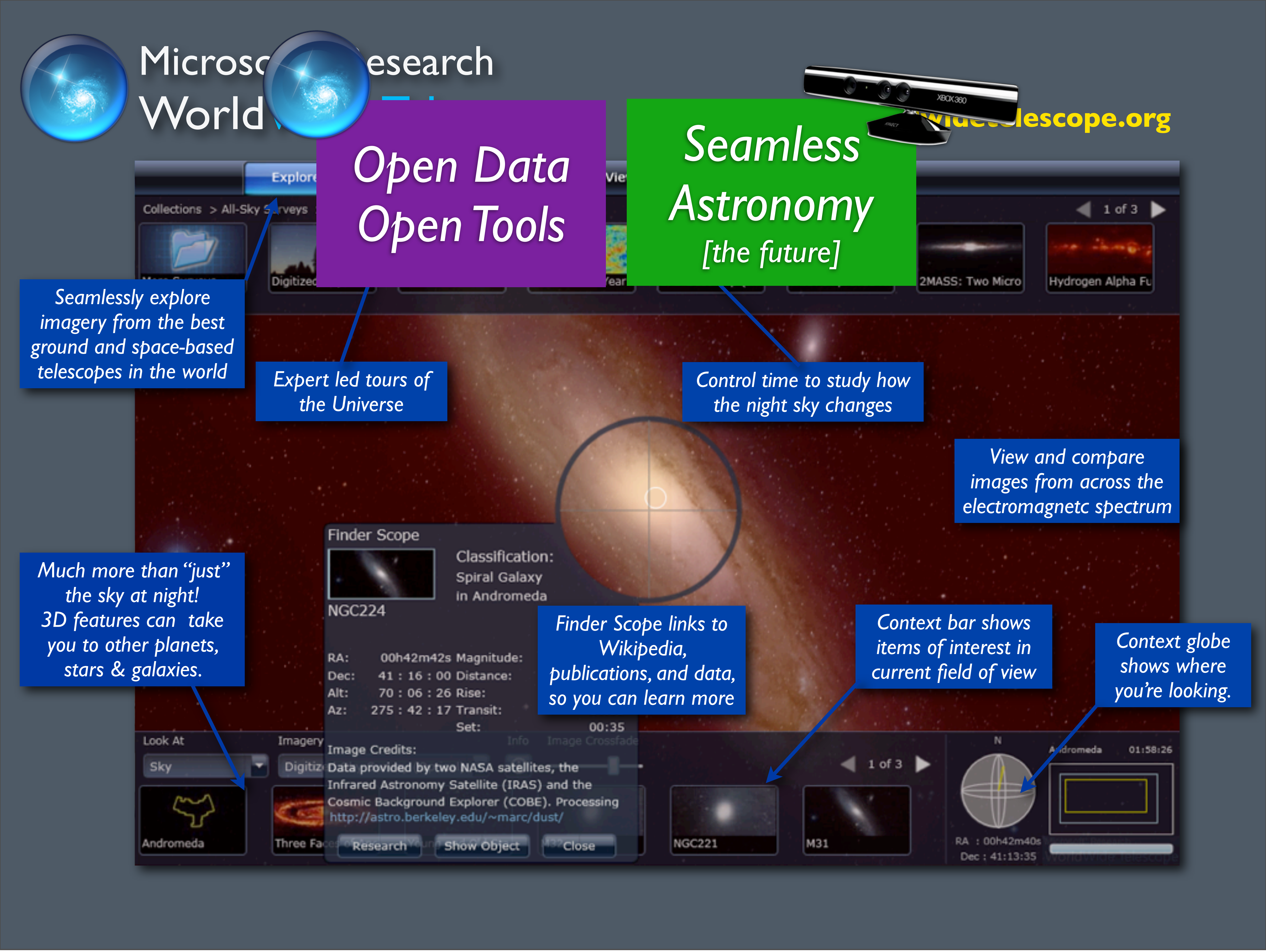


**thesis of Tom Rice, '12, awarded Hoopes Prize*

Value of “Linked Views” from Harvard Undergrad Thesis* Work

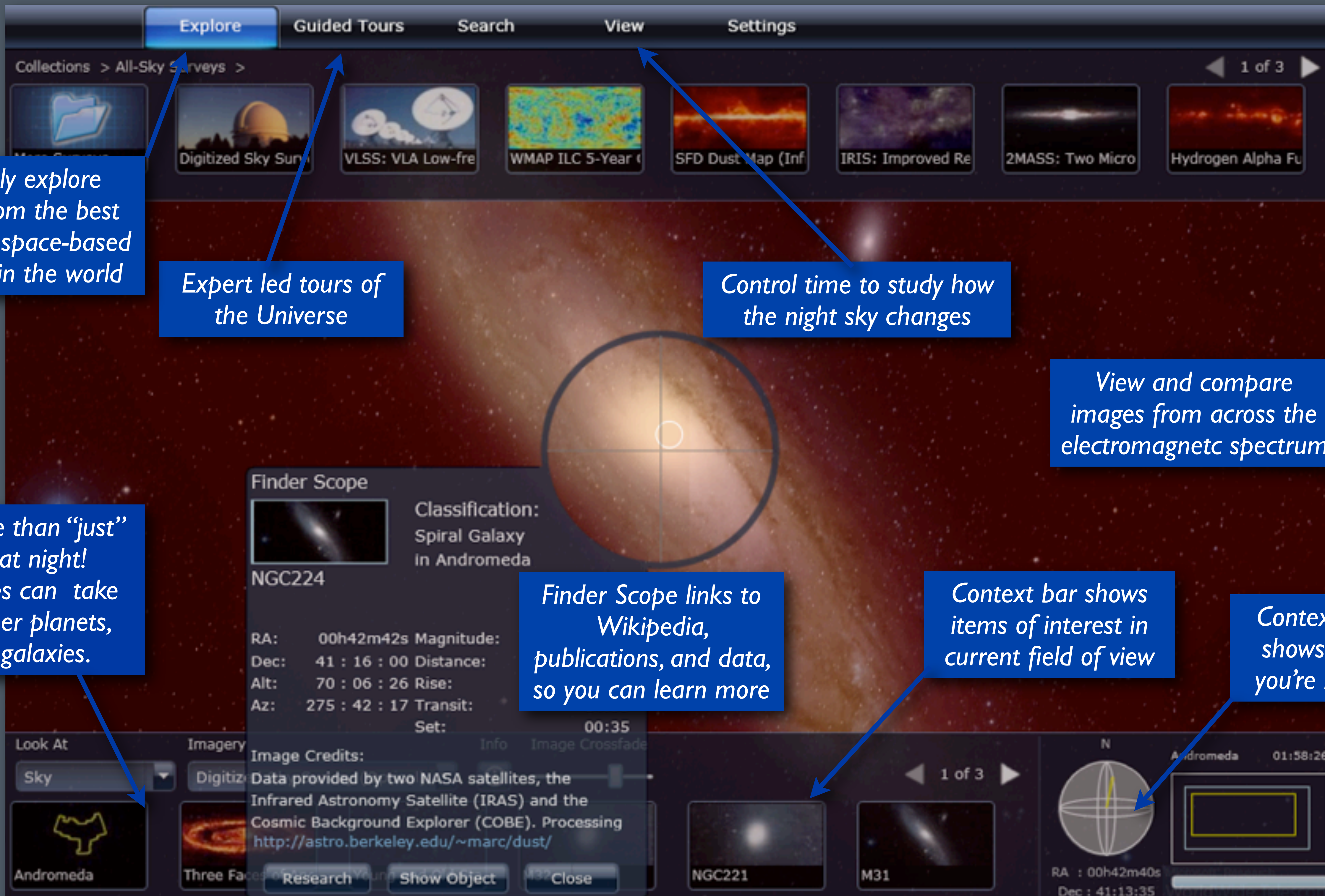


thesis of Tom Rice, ’12, awarded Hoopes Prize


Microsoft Research
World Wide Telescope
worldwidetelescope.org
Open Data
Open Tools
Seamless
Astronomy
[the future]
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.
Explore
Collections > All-Sky Surveys
1 of 3
2MASS: Two Micro
Hydrogen Alpha Fu
Finder Scope
Classification:
Spiral Galaxy
in Andromeda
NGC224
RA: 00h42m42s Magnitude:
Dec: 41 : 16 : 00 Distance:
Alt: 70 : 06 : 26 Rise:
Az: 275 : 42 : 17 Transit:
Set: 00:35
Image Credits:
Data provided by two NASA satellites, the Infrared Astronomy Satellite (IRAS) and the Cosmic Background Explorer (COBE). Processing
http://astro.berkeley.edu/~marc/dust/
Research
Show Object
Close
Look At
Sky
Imagery
Andromeda
NGC221
M31
RA : 00h42m40s
Dec : 41:13:35
01:58:26

Microsoft® Research WorldWide Telescope

Experience WWT at **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.

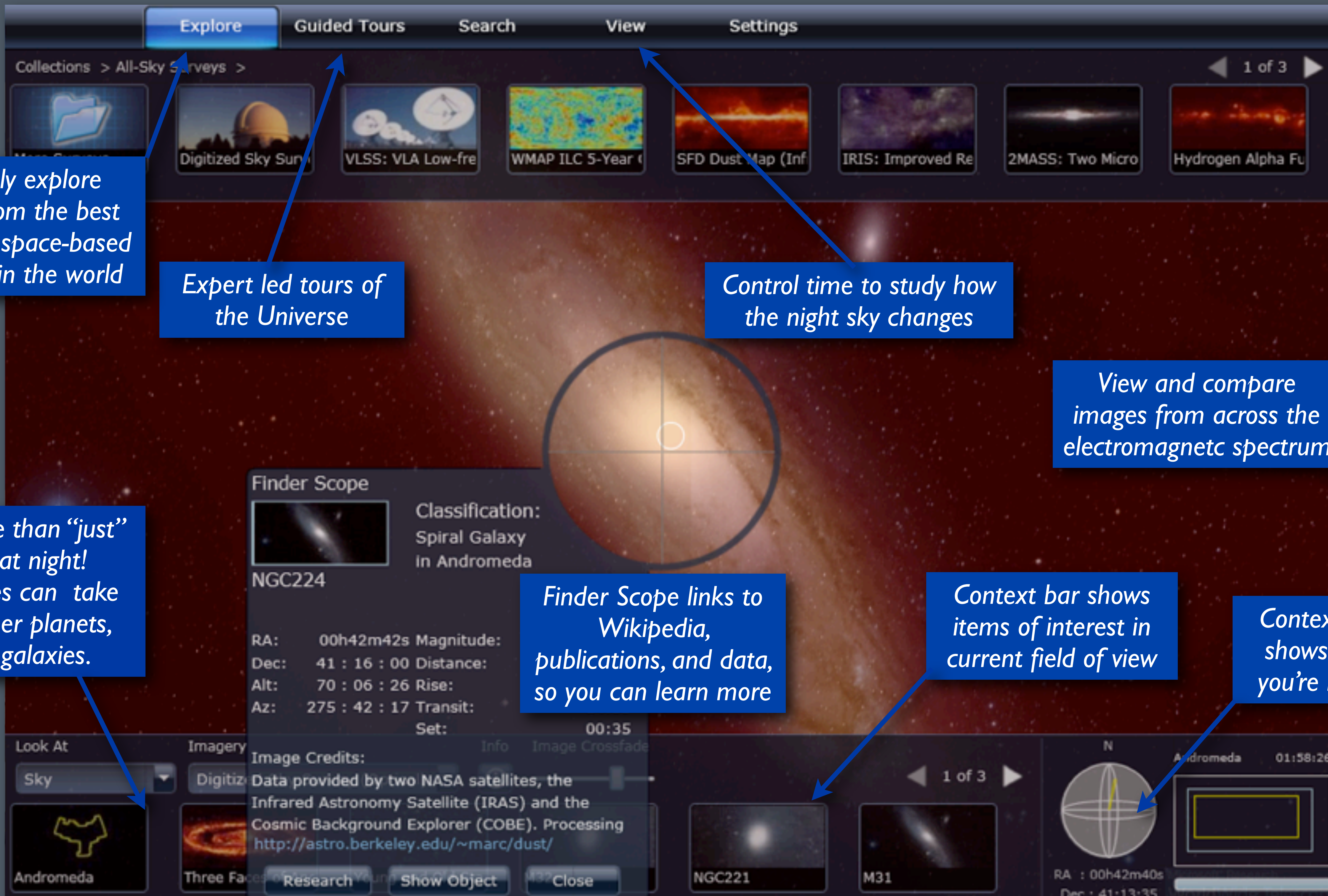
Finder Scope links to Wikipedia, publications, and data, so you can learn more

Context bar shows items of interest in current field of view

Context globe shows where you're looking.

Microsoft® Research WorldWide Telescope

Experience WWT at **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.

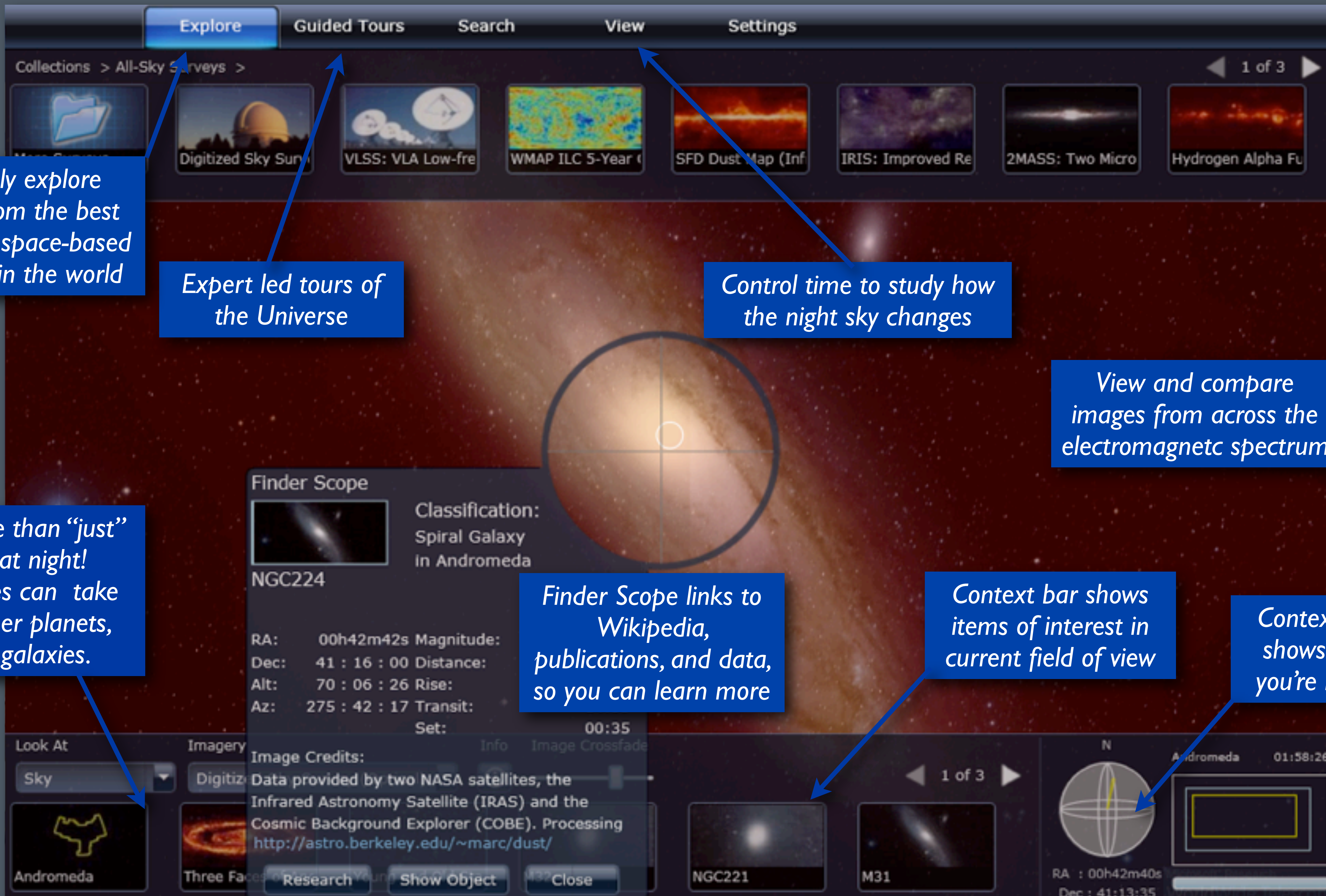
Finder Scope links to Wikipedia, publications, and data, so you can learn more

Context bar shows items of interest in current field of view

Context globe shows where you're looking.

Microsoft® Research WorldWide Telescope

Experience WWT at **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

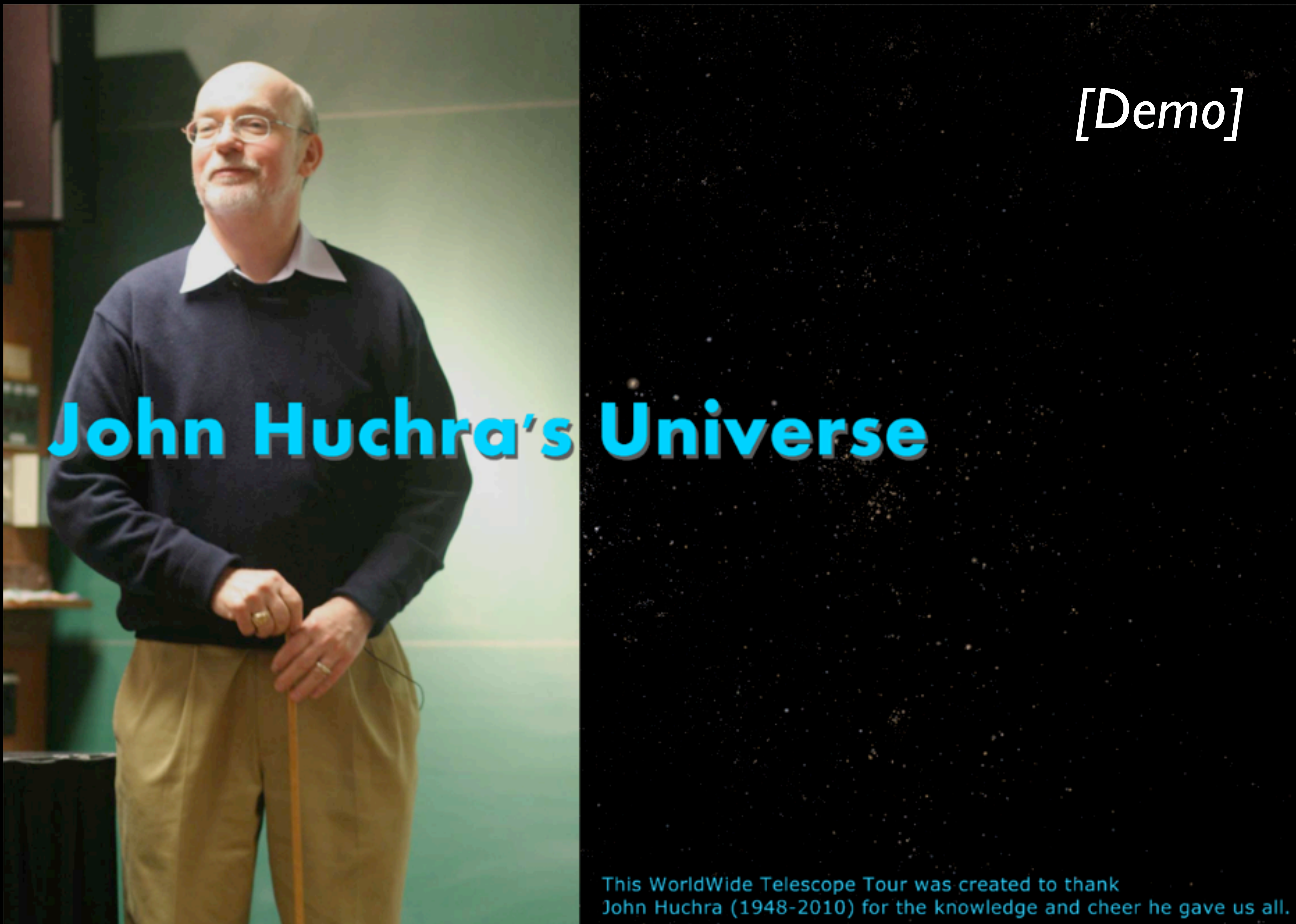
View and compare images from across the electromagnetc spectrum

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.



also available on *YouTube* *(search “John Huchra’s Universe”)*

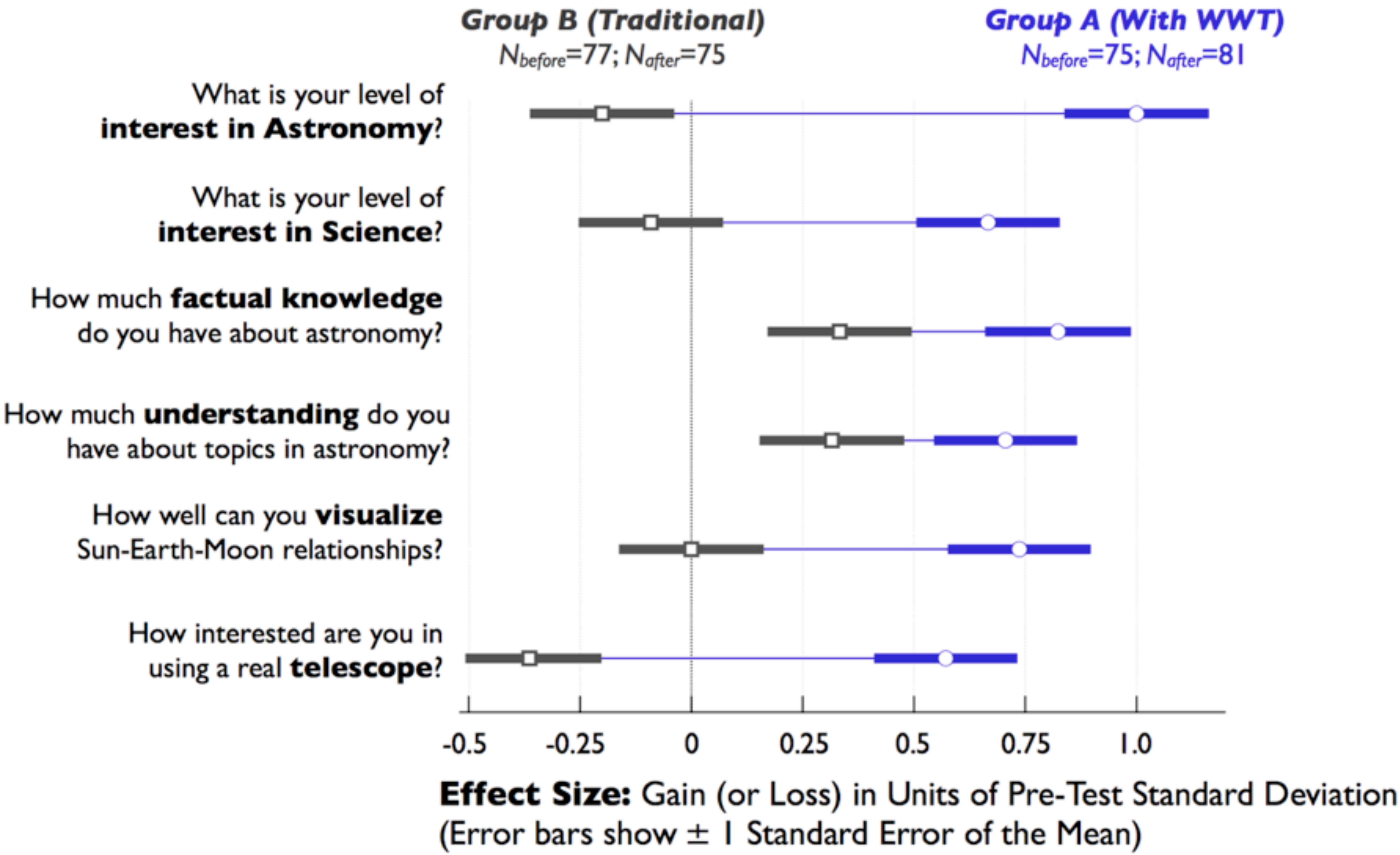
The WorldWide Telescope Ambassadors Program

Alyssa **Goodman** & Patricia **Udomprasert**
Harvard-Smithsonian Center for Astrophysics

Curtis Wong & Jonathan Fay
Microsoft Research

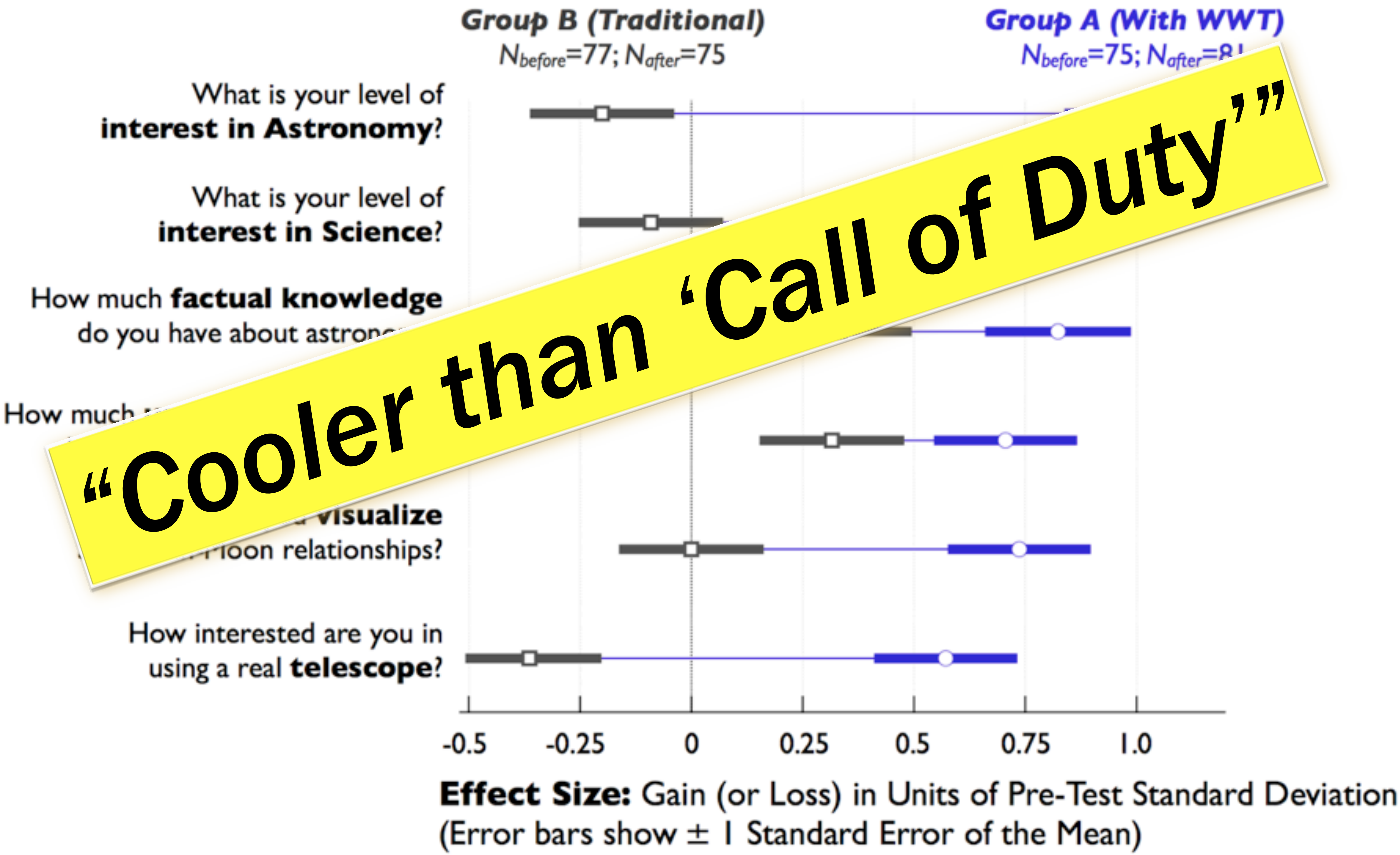
Gains in Student Interest and Understanding

(“Traditional Way” vs “WWT Way”)



Gains in Student Interest and Understanding

(“Traditional Way” vs “WWT Way”)



RVO

Astro Grid

Aus-VO

VObs.it

vo-i

CVO

SVO
Spanish Virtual Observatory

BRAVO

ArVO
Armenian Virtual Observatory

NVO
NATIONAL VIRTUAL OBSERVATORY

IVOA

China-VO

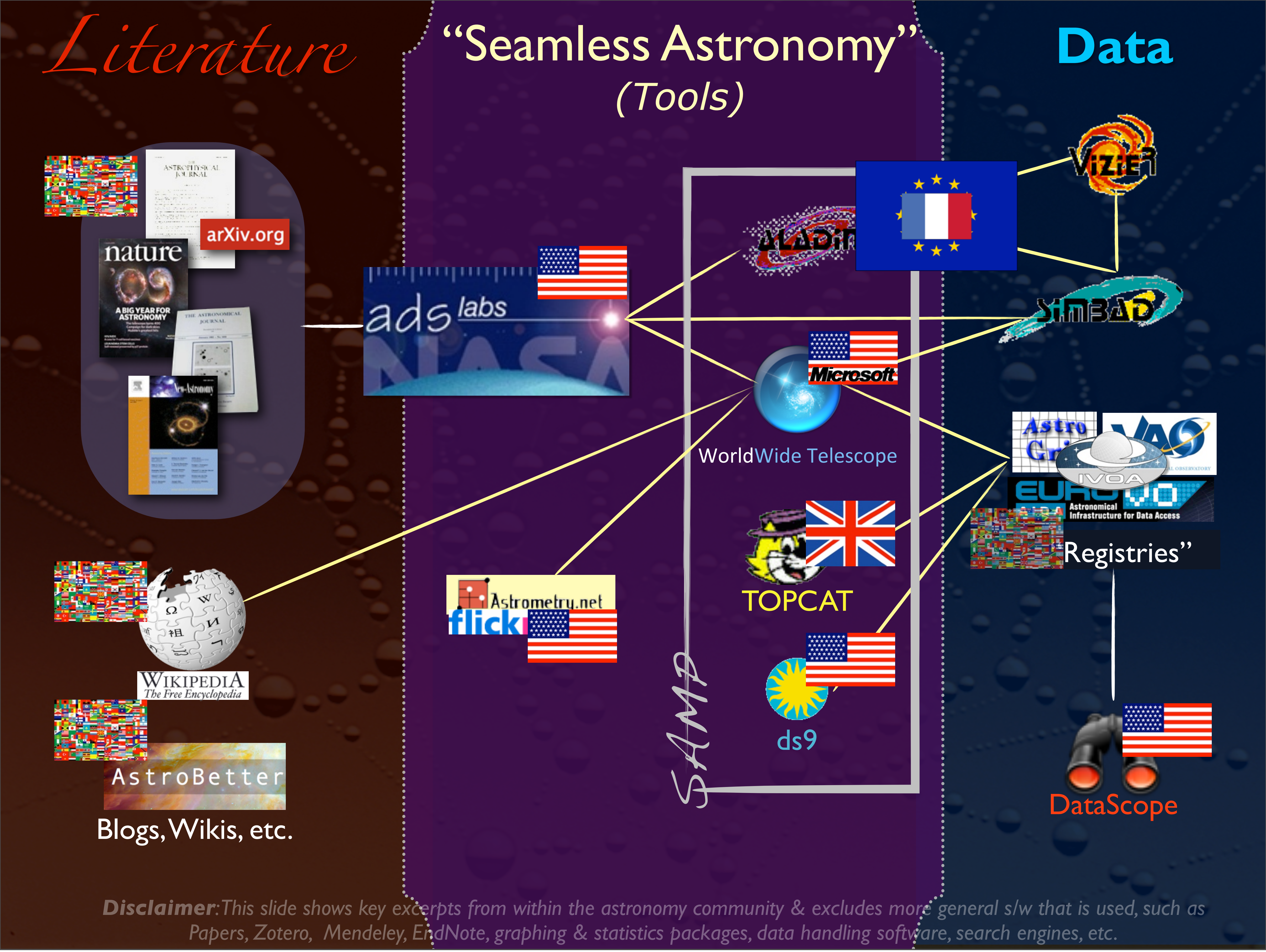
EURO VO

JVO

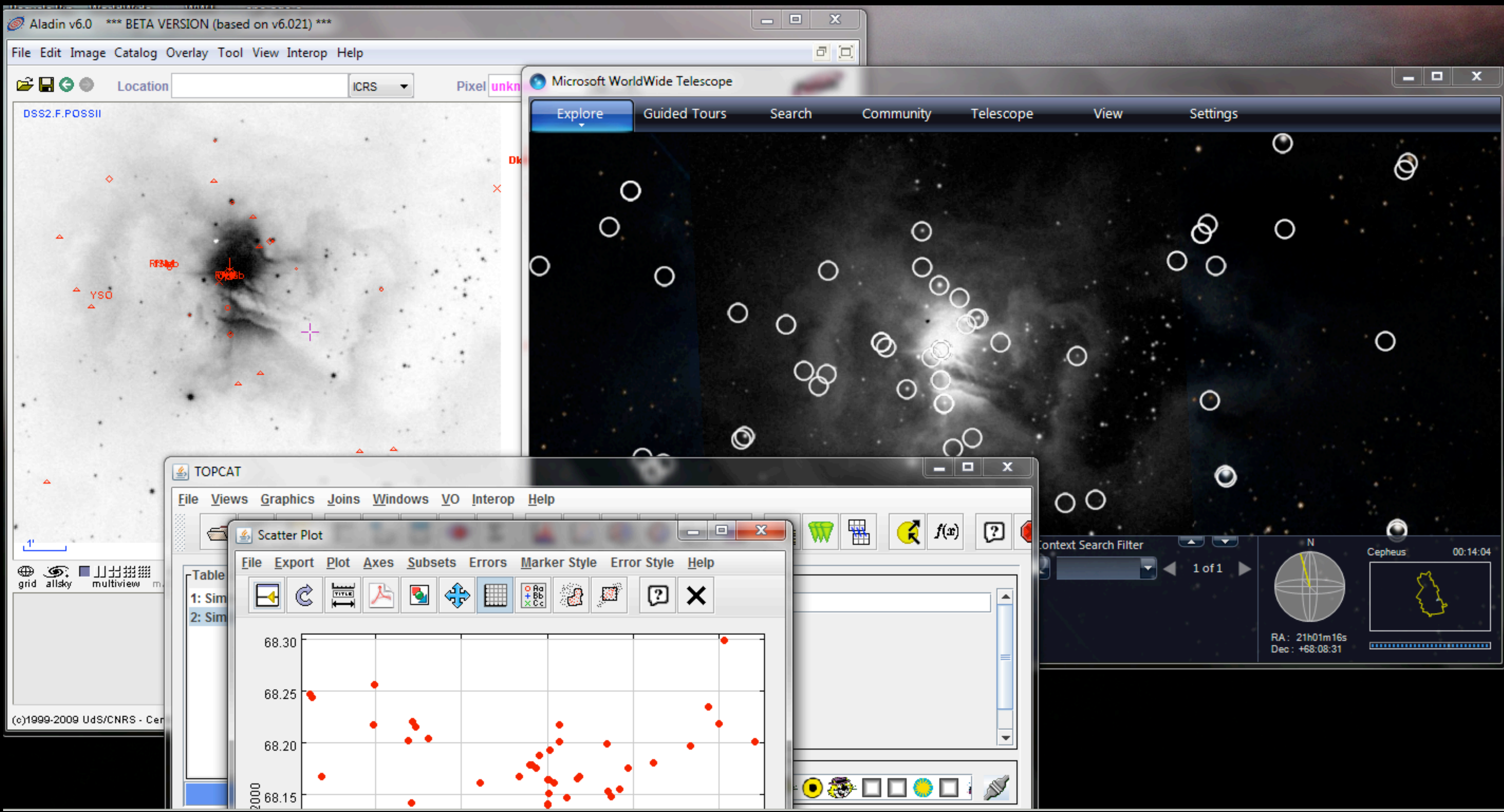
HVO
Hungarian Virtual Observatory

GERMAN ASTROPHYSICAL
GAVO
VIRTUAL OBSERVATORY

France

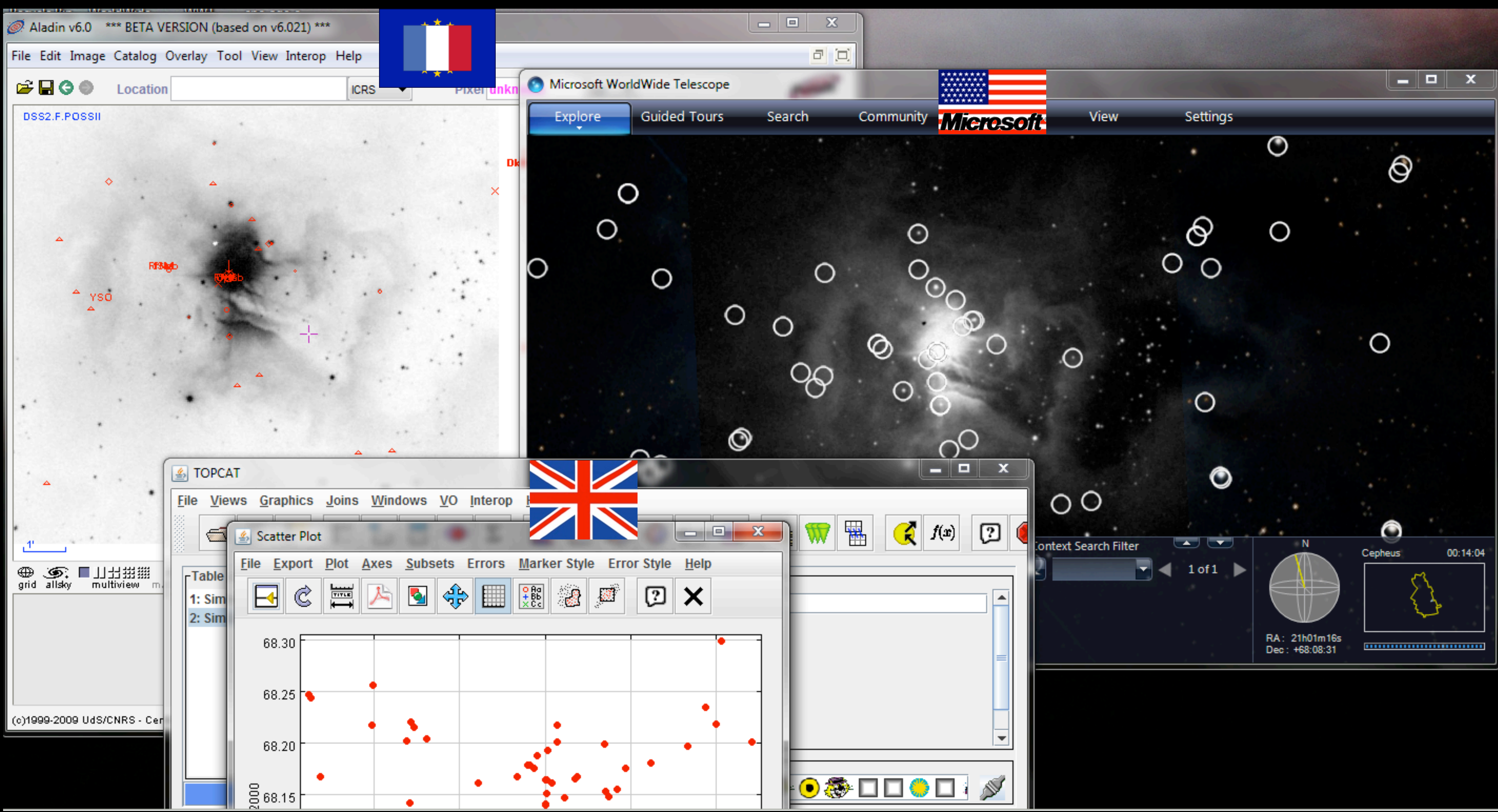

Literature
“Seamless Astronomy”
(Tools)
Data
arXiv.org
nature
ads labs
NASA
ALADIN
VizieR
SIMBAD
Microsoft
WorldWide Telescope
AstroGrid
IVOA
EURO VO
Astronomical Infrastructure for Data Access
Registries”
TOPCAT
Astrometry.net
flickr
ds9
SAMP
WIKIPEDIA
The Free Encyclopedia
AstroBetter
Blogs, Wikis, etc.
DataScope
Disclaimer: This slide shows key excerpts from within the astronomy community & excludes more general s/w that is used, such as Papers, Zotero, Mendeley, EndNote, graphing & statistics packages, data handling software, search engines, etc.

SAMP
(Simple Application Messaging Protocol)



link to 12/2010 IVOA recommendation

SAMP
(Simple Application Messaging Protocol)



link to 12/2010 IVOA recommendation

SEAMLESS ASTRONOMY

Linking scientific data, publications, and communities

ABOUT PROJECTS PEOPLE RESOURCES DATAVERSE

SHARE

SEAMLESS ASTRONOMY

About

The **Seamless Astronomy Group** at the **Harvard-Smithsonian Center for Astrophysics** brings together astronomers, computer scientists, information scientists, librarians and visualization experts involved in the development of tools and systems to study and enable the next generation of **online astronomical research**.

Current projects include research on the development of systems that seamlessly integrate scientific data and literature, the semantic interlinking and annotation of scientific resources, the study of the impact of social media and networking sites on scientific dissemination, and the analysis and visualization of astronomical research communities. Visit our project page to find out more.

Latest Announcements

Introducing the Astronomy Dataverse

Latest Feed Items

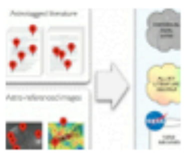
@rahuldave there is a writeboard with my notes... More at next #seamlessastronomy next week.

Thanks to @astrobites and @astroknight06 for great summary http://t.co/jWWFT0CD of our High-D Data Viz work! #ivoa #seamlessastronomy

SEAMLESS ASTRONOMY

Projects

Seamless integration of scientific data and literature
Astronomical data artifacts and publications exist in disjointed repositories. The conceptual relationship that links data and publications is rarely made explicit. In collaboration with **ADS** and **ADSlabs**, and through our work in conjunction with the Institute for Quantitative Social Science (IQSS), we are working on developing a platform that allows data and literature to be seamlessly integrated, interlinked, mutually discoverable.



ADS All-SKy Survey (ADSASS)
The ADS All-Sky Survey (ADSASS) is an ongoing effort aimed at turning the NASA Astrophysics Data System (ADS), widely known for its unrivaled value as a literature resource for astronomers, into a data resource. The ADS is not a data repository per se, but it implicitly contains valuable holdings of astronomical data, in the form of images, tables and object references contained within articles. The objective of the ADSASS effort is to extract these data and make them discoverable and available through existing data viewers. The resulting ADSASS data layer promises to greatly enhance workflows and enable new research by tying astronomical literature and data assets into one resource. More information can be found on this conference paper.

Astronomy Dataverse
Astronomers use, peruse and produce vast amounts of scientific data. Making these data publicly available is important because it supports the reproducibility of results, and ensures their long term preservation and reuse. While raw astronomical data are normally stored and made public available via large-scale archives, reduced data are often left out entirely from both astronomical archives and related publications.

In a pilot study in 2011, we are evaluating the Dataverse, an open data archive hosted by Harvard University and managed by the Institute for Quantitative Social Science (IQSS), as a project-based repository for the storage, access, and citation of reduced astronomical data. We have interviewed a set of 10 astronomers about their needs, and the **prototype CfA Dataverse** is now online.

WorldWide Telescope (WWT)
WorldWide Telescope provides a rich contextual visualization environment for astronomical data. Our group collaborates with the WWT Team at Microsoft Research both to enrich WWT for use in research as well as in teaching. On the research end, we seek to integrate WWT "Seamlessly" with VAO-sponsored projects, as well as with ADS Labs. On the teaching end, we founded and now run the **WorldWide Telescope Ambassadors** outreach effort.

Viz-e-Lab
Established in 2011, the **Viz-e-Lab** was established as a testing ground for new software efforts in visualization and e-Science at the CfA. Seamless Astronommy projects are **piloted and tested on users** in this space, located on the third floor of the 160 Concord Avenue building of the CfA. The lab is used to test new hardware--primarily input devices--as well as new software. At present, two main focii are the development of sophisticated tools "linked view" visualization of high-dimensional data, and the integration of **WorldWide Telescope** into research and teaching paradigms.

Study of the impact of social media and networking sites on scientific dissemination
Astronomers, and more broadly, the scientific community, are increasingly using blogging, micro-blogging, and other social media for both discovering and disseminating scientific knowlegde. We are exploring several avenues for studying the impact of Twitter and other social networking sites on scientific readership.

Network analysis and visualization of astronomical research communities
We use network analytic techniques to mine the astronomical bibgliographic archives and detect disciplinary and geographical clusters, and communities of practices of scientists. A network visualization of co-authorship networks in Physics and Astronomy is being implemented on the ADSlabs platform and is currently available. Try this out on ADS Labs by doing a search and then choosing "View as Author Network."

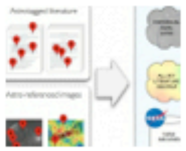
Data citation practices in Astronomy
How do astronomers cite scientific data? Are astronomical reduced data included in scholarly papers as supplemental material? Are cited data discoverable and reusable? We are performing link and content analyses of bibliographic repositories in astronomy to understand whether astronomical data used for the publication of scientific research can be discovered, accessed, and reused. On ADS Labs, choose "View as Paper Network" after a search to get a feel for the data available.

Semantic description and annotation of scientific resources
RDF store and facets, and links to semantic ADS.

SEAMLESS ASTRONOMY

Projects

ADS Labs



ADS All-SKy Survey (ADSASS)

ADS All Sky Survey

can be found on this conference paper.

Astronomy Dataverse

Astronomy Dataverse

hosted by Harvard University and managed by the Institute for Quantitative Social Science (IQSS), as a project-based repository for the storage, access, and citation of reduced astronomical data. We have interviewed a set of 10 astronomers about their needs, and the prototype CfA Dataverse is now online.

WorldWide Telescope

Viz-e-Lab

High-D Visualization

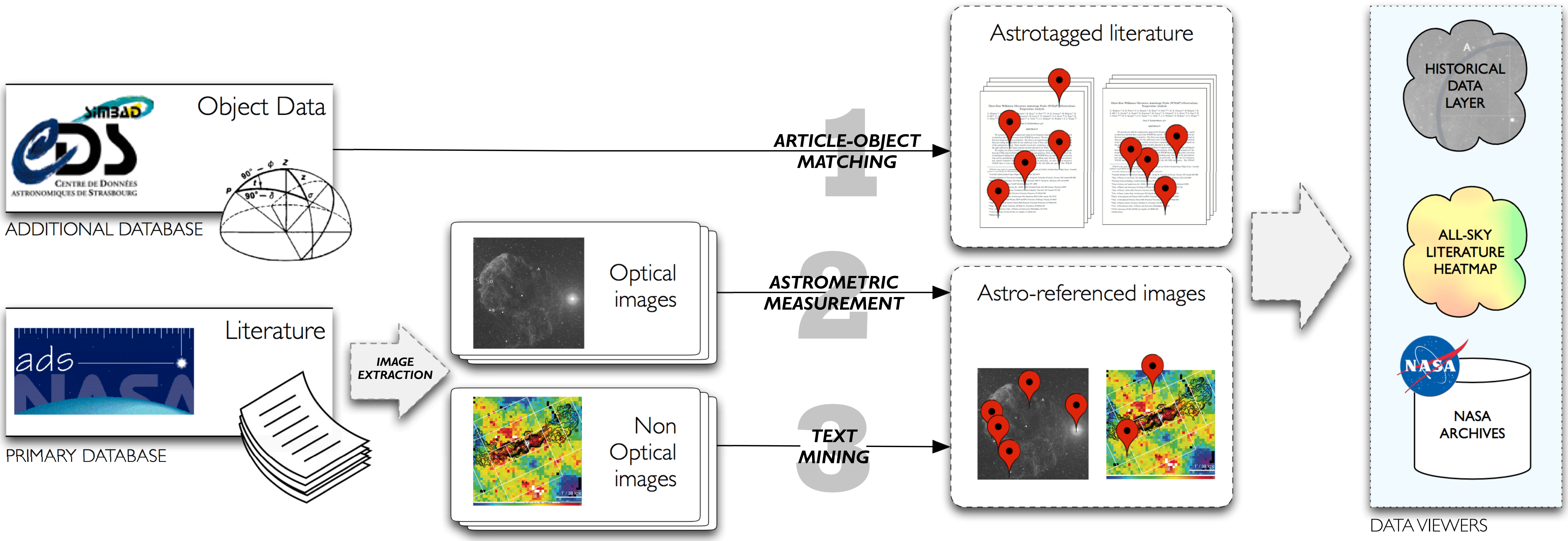
high-dimensional data, and the integration of WorldWide Telescope into research and teaching paradigms.

Social Networks

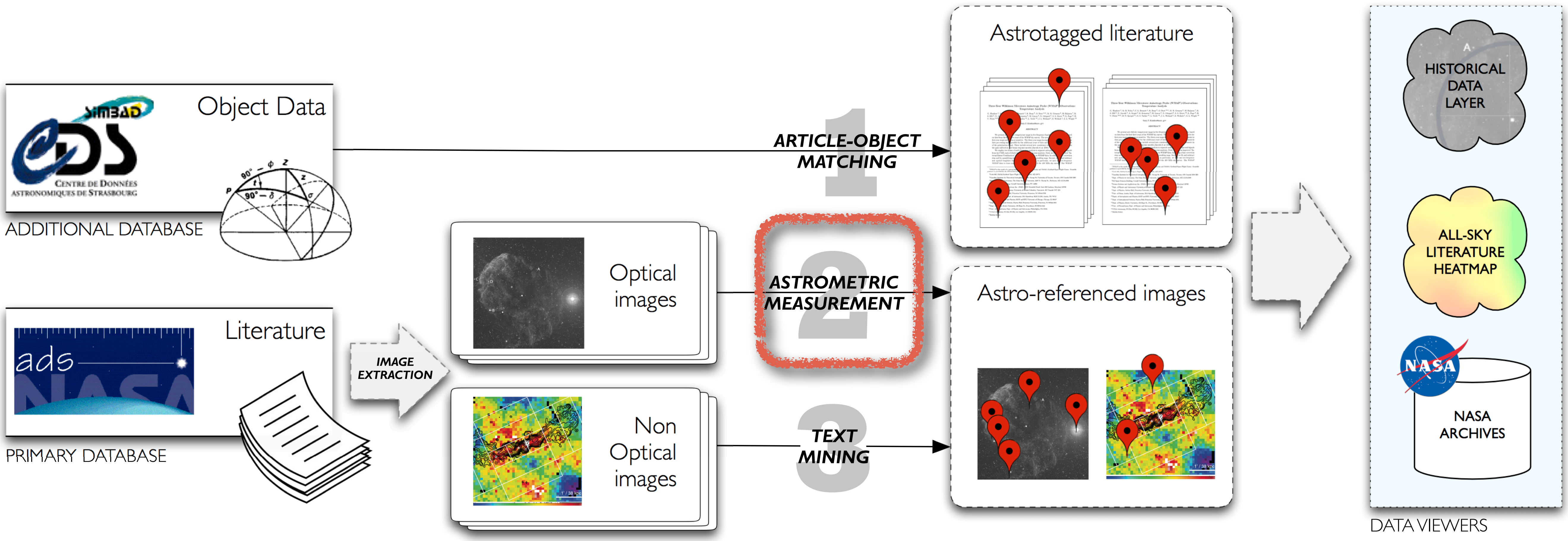
Collaboration Networks

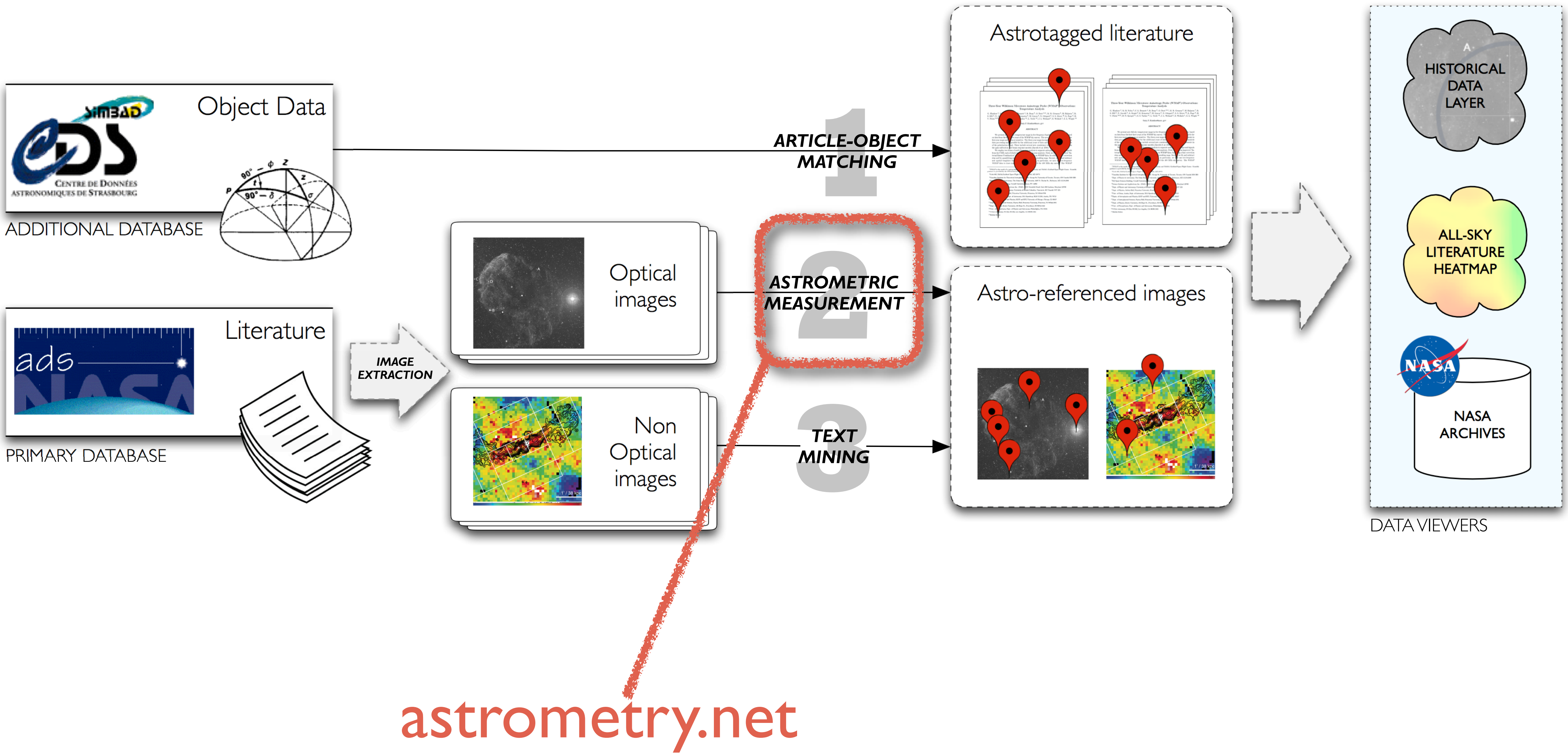
Data Citation

Semantic Search



Object Data
SIMBAD
CDS
CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG
ADDITIONAL DATABASE
Literature
ads
NASA
PRIMARY DATABASE
IMAGE EXTRACTION
Optical images
Non Optical images
1
ARTICLE-OBJECT MATCHING
2
ASTROMETRIC MEASUREMENT
3
TEXT MINING
Astrotagged literature
Astro-referenced images
HISTORICAL DATA LAYER
ALL-SKY LITERATURE HEATMAP
NASA
NASA ARCHIVES
DATA VIEWERS

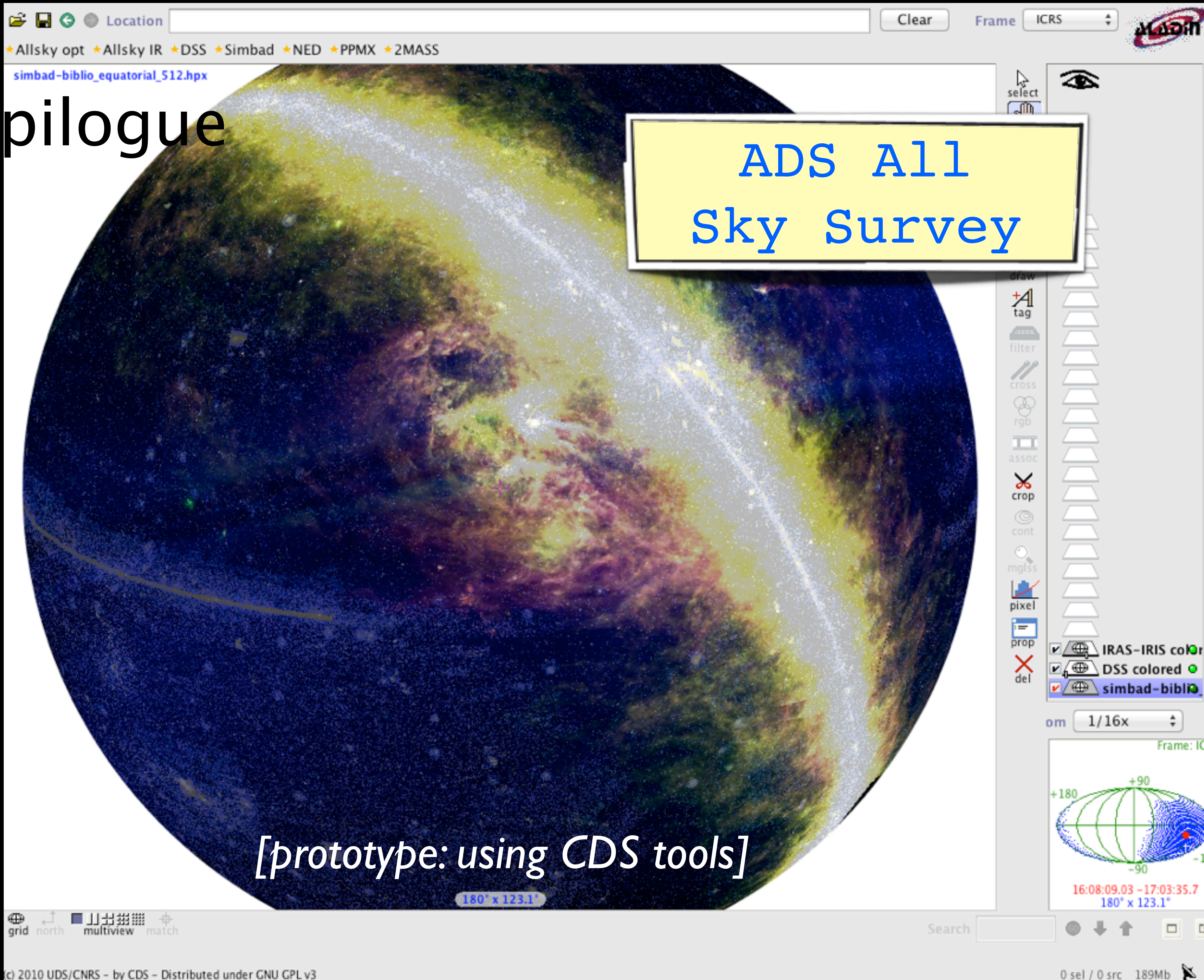

Object Data
ADDITIONAL DATABASE
Literature
PRIMARY DATABASE
IMAGE EXTRACTION
Optical images
Non Optical images
1
ARTICLE-OBJECT MATCHING
2
ASTROMETRIC MEASUREMENT
3
TEXT MINING
Astrotagged literature
Astro-referenced images
HISTORICAL DATA LAYER
ALL-SKY LITERATURE HEATMAP
NASA ARCHIVES
DATA VIEWERS



Object Data
CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG
ADDITIONAL DATABASE
Literature
ads
PRIMARY DATABASE
IMAGE EXTRACTION
Optical images
Non Optical images
ARTICLE-OBJECT MATCHING
ASTROMETRIC MEASUREMENT
TEXT MINING
Astrotagged literature
Astro-referenced images
HISTORICAL DATA LAYER
ALL-SKY LITERATURE HEATMAP
NASA
NASA ARCHIVES
DATA VIEWERS
astrometry.net

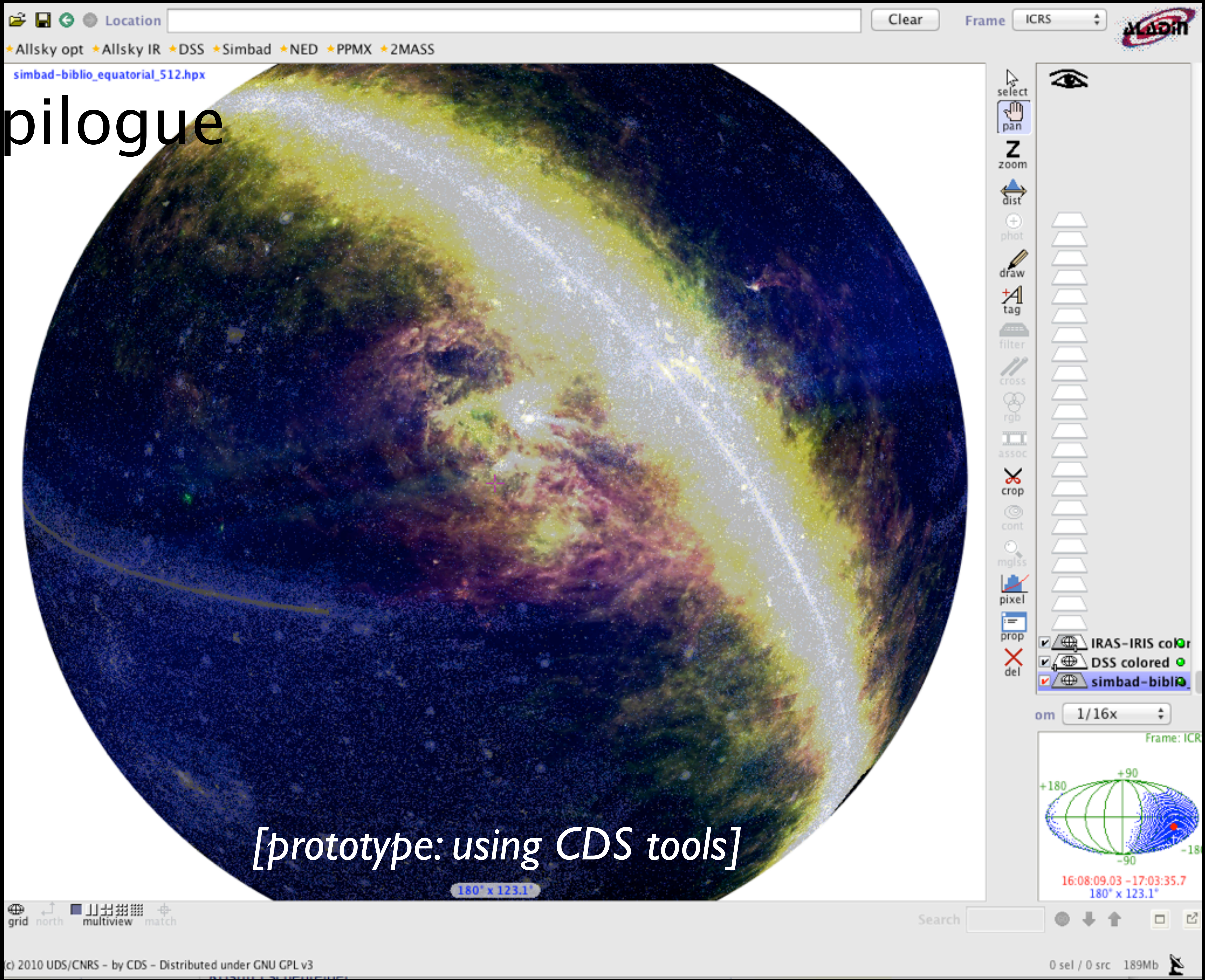
pilogue

ADS All Sky Survey



[prototype: using CDS tools]

pilogue

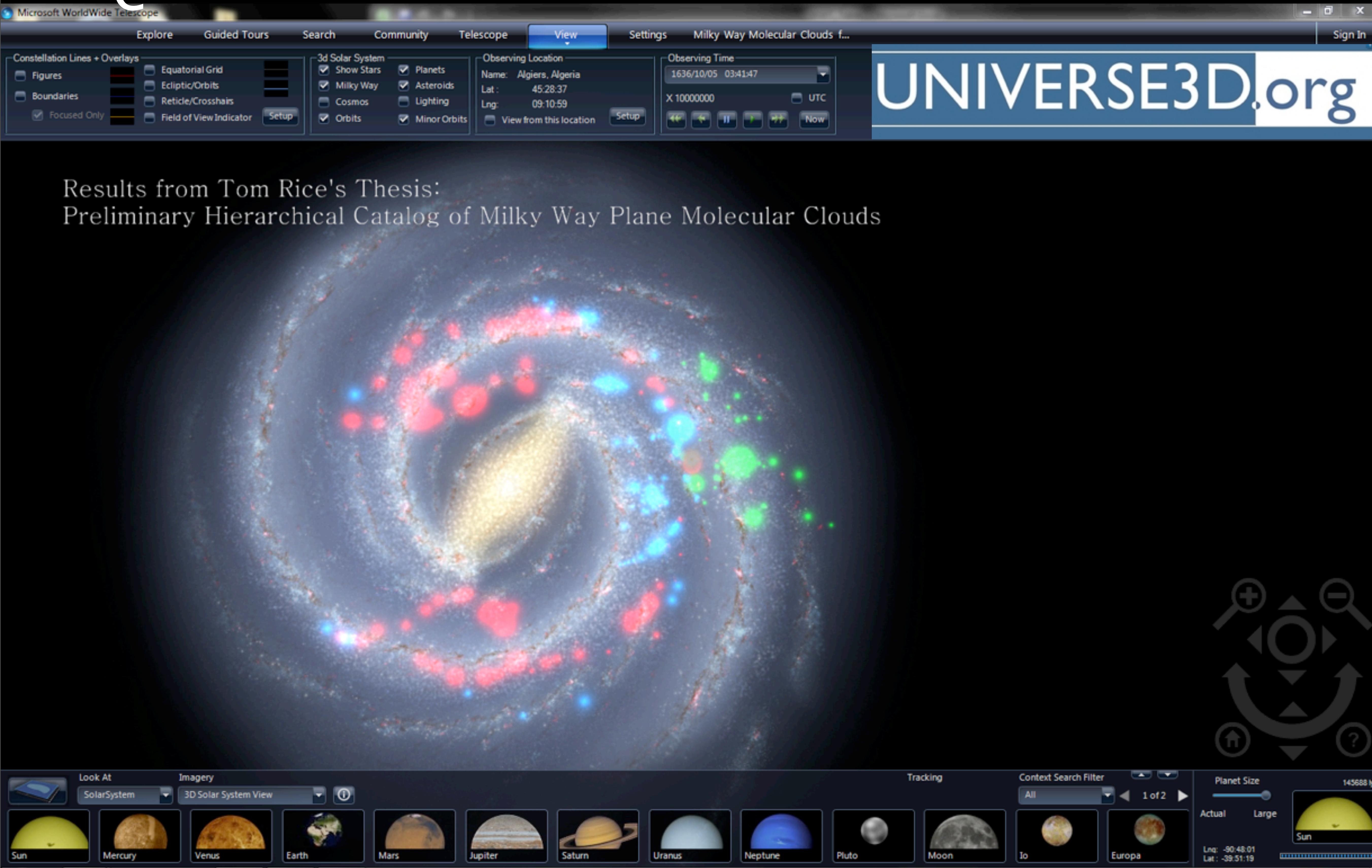


[prototype: using CDS tools]

pilogue


EN · The Milky Way Project is part of the ZOONIVERSE
...just like SOLAR STORMWATCH
THE MILKY WAY PROJECT
FOLLOW US ON TWITTER
VISIT THE BLOG
MILKY WAY TALK
HOME
TAKE PART
ABOUT
TUTORIAL
LOG IN
GALACTOMETER™
DRAW BUBBLES
WELCOME
The Milky Way Project aims to sort and measure our galaxy, the Milky Way. Initially we're asking you to help us find and draw bubbles in beautiful infrared data from the Spitzer Space Telescope.
Understanding the cold, dusty material that we see in these images, helps scientists to learn how stars form and how our galaxy changes and evolves with time.
Click here to see the full tutorial or browse the site to find out more about the science behind the Milky Way Project.
YOU CAN NOW SEE HOW CLOSE WE ARE TO 1,000,000 DRAWINGS AT HTTP://WWW.MILKYWAYPROJECT.ORG/G... 12 DAYS AGO
194,943 IMAGES SERVED · 252,562 BUBBLES DRAWN · 24,234 POSSIBLE STAR CLUSTERS · 8,978 CANDIATE GALAXIES · 597,054 OTHER OBJECTS
© COPYRIGHT 2010 ZOONIVERSE

Epilogue


Microsoft WorldWide Telescope
Explore
Guided Tours
Search
Community
Telescope
View
Settings
Milky Way Molecular Clouds f...
Sign In
Constellation Lines + Overlays
Figures
Boundaries
Focused Only
Equatorial Grid
Ecliptic/Orbits
Reticle/Crosshairs
Field of View Indicator
Setup
3d Solar System
Show Stars
Milky Way
Cosmos
Orbits
Planets
Asteroids
Lighting
Minor Orbits
Observing Location
Name: Algiers, Algeria
Lat: 45:28:37
Lng: 09:10:59
View from this location
Setup
Observing Time
1636/10/05 03:41:47
X 10000000
UTC
Now
UNIVERSE3D.org
Results from Tom Rice's Thesis:
Preliminary Hierarchical Catalog of Milky Way Plane Molecular Clouds
Look At
SolarSystem
Imagery
3D Solar System View
Tracking
Context Search Filter
All
1 of 2
Planet Size
Actual
Large
Sun
Mercury
Venus
Earth
Mars
Jupiter
Saturn
Uranus
Neptune
Pluto
Moon
Io
Europa
Sun

pilogue

UNIVERSE3D.org

Agoodman My talk My preferences My watchlist My contributions Log out

Page Discussion

Read Edit View history Go Search

What is Universe3D.org?

Navigation

- Home
- 3D Viewers
- Datasets
- Images
- Publications & Presentations
- People

More

- Calendar
- Announcements
- Help

Toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

The intention of Universe3D.org is to host links to web content that enable the enhancement of our three-dimensional view of the Universe.

Recently added Dataset

SLOAN Digital Sky Survey The Sloan Digital Sky Survey or SDSS is a major multi-filter imaging and spectroscopic redshift survey using a dedicated 2.5-m wide-angle optical telescope at Apache Point Observatory in New Mexico, United States. The main galaxy sample has a median redshift of z = 0.1; there are redshifts for luminous red galaxies as far as z = 0.7, and for quasars as far as z = 5; and the imaging survey has been involved in the detection of quasars beyond a redshift z = 6.

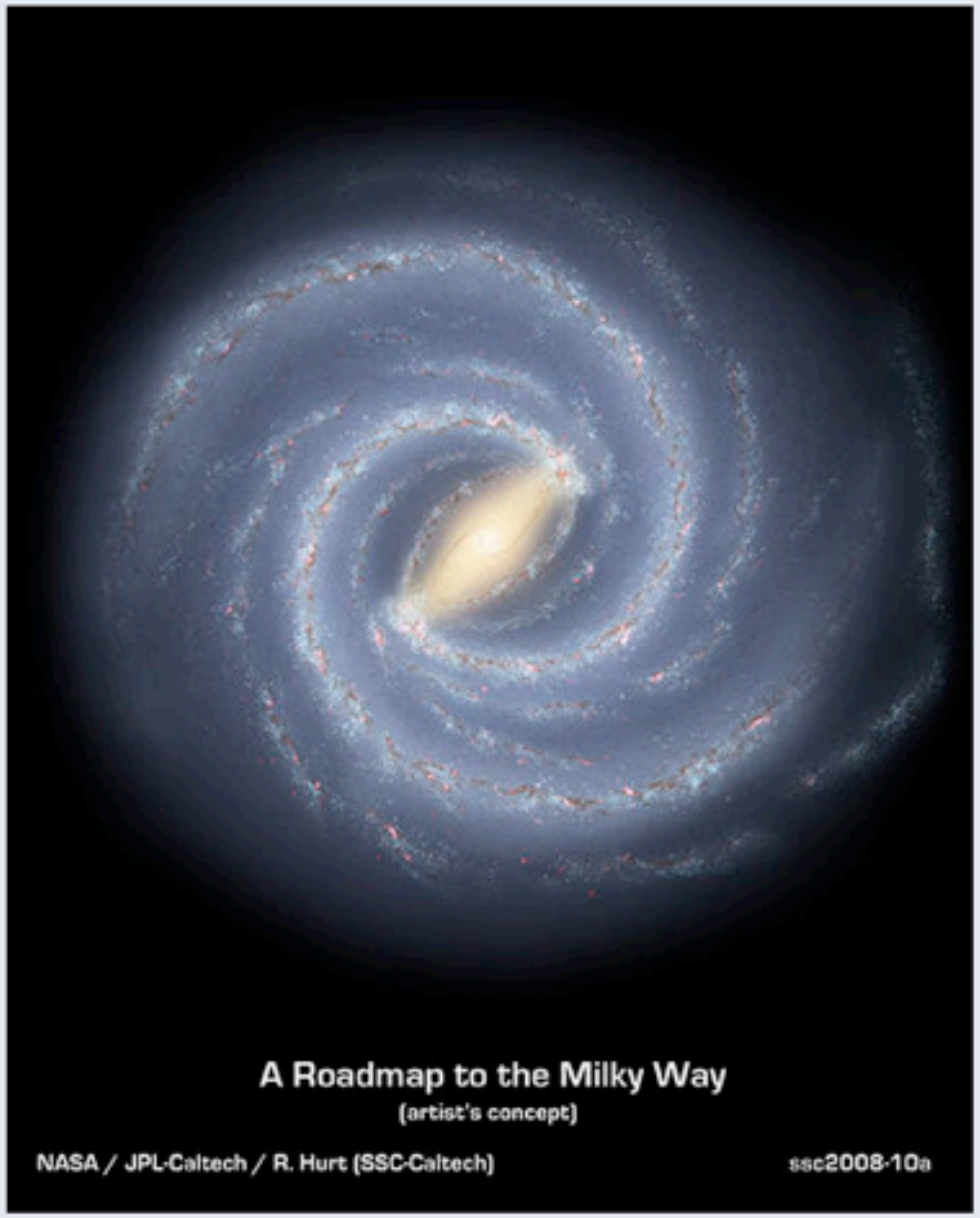
Astronomy News

- *June 26, 2012:* Astronomers use supercomputer to explore role of dark matter in galaxy formation
- *June 25, 2012:* Moon to pass by Mars tonight
- *June 24, 2012:* Astronomers find planets so close they 'see' each other in night sky
- *June 14, 2012:* Huge Asteroid to fly by Earth
- *June 13, 2012:* Astronomers may have discovered the oldest galaxy in the Universe
- *June 5, 2012:* Last Transit of Venus for the 21st century

Announcements

- *July 05, 2012:* Website moved to the URL universe3d.org!
- *June 11, 2012:* Website moved to MediaWiki!
- *December 5, 2011:* Site established!
  To make good on Alyssa Goodman's promise at the "Milky Way 2011" meeing held in Rome this past September, the site "universe3d.org" has been established. By 2012, it will be populated with links to existing data

The Milky Way



2002

The World-Wide Telescope, an Archetype for Online Science

Jim Gray
Microsoft Research
Gray@Microsoft.com

Alex Szalay
The Johns Hopkins University
Szalay@jhu.edu

Abstract Most scientific data will never be directly examined by scientists; rather it will be put into online databases where it will be analyzed and summarized by computer programs. Scientists increasingly see their instruments through online scientific archives and analysis tools, rather than examining the raw data. Today this analysis is primarily driven by scientists asking queries, but scientific archives are becoming active databases that self-organize and recognize interesting and anomalous facts as data arrives. In some fields, data from many different archives can be cross-correlated to produce new insights. Astronomy presents an excellent example of these trends; and, federating Astronomy archives presents interesting challenges for computer scientists.

Introduction
Computational Science is a new branch of most disciplines. A thousand years ago, science was primarily *empirical*. Over the last 500 years each discipline has grown a *theoretical* component. Theoretical models often motivate experiments and generalize our understanding. Today most disciplines have both empirical and theoretical branches. In the last 50 years, most disciplines have grown a third, *computational* branch (e.g. empirical, theoretical, and computational ecology, or physics, or linguistics.)

statistics among sets of data points in a metric space. Pair-algorithms on N points scale as N^2. If the data increase a thousand fold, the work and time can grow by a factor of a million. Many clustering algorithms scale even worse. These algorithms are infeasible for terabyte-scale datasets.

The new online science needs new data mining algorithms that use near-linear processing, storage, and bandwidth, and that can be executed in parallel. Unlike current algorithms that give exact answers, these algorithms will likely be heuristic and give approximate answers [Connolly, Szapudi].

Astronomy as an Archetype for Online Science
Astronomy exemplifies these phenomena. For thousands of years astronomy was primary empirical with few theoretical models. Theoretical astronomy began with Kepler is now co-equal with observation. Astronomy was early to adopt computational techniques to model stellar and galactic formation and celestial mechanics. Today, simulation is an important part of the field – producing new science, and solidifying our grasp of existing theories.

Astronomers are building telescopes that produce terabytes of data each year -- soon terabytes per night. In the old

2012

Authorea

Get your research done. Collaboratively. On the web.

Are your research data, images, software, and manuscripts scattered around emails, Dropbox folders, hard drives, servers, and data archives? **We are fixing that!**

We are building ***Authorea***, an online platform for the collaborative authorship of research projects. It will let you publish, share, organize, version control, and source control all the components of your research.

Authorea is an incubator initiative of **Harvard University** and the Harvard-Smithsonian Center for Astrophysics.

Enter your e-mail address and we'll keep you in the loop.

email@address.com GO

powered by launchrock

[Pepe et al.... coming soon]