

The background of the slide is a vibrant cosmic image featuring a nebula with swirling clouds of gas in shades of blue, green, and yellow, punctuated by numerous bright stars and distant galaxies.

Digital Data Preservation and Curation: A Collaboration Among Libraries, Publishers, and the Virtual Observatory

*A pilot project aimed at preserving, curating, and enabling access
to digital data and associated electronic journals content.*

Teresa Ehling, Cornell University

Robert Hanisch, Space Telescope Science Institute

Julie Steffen, University of Chicago Press

*Sayeed Choudhury, Tim DiLauro, Alex Szalay, and Ethan Vishniac,
The Johns Hopkins University*

Robert Milkey, American Astronomical Society

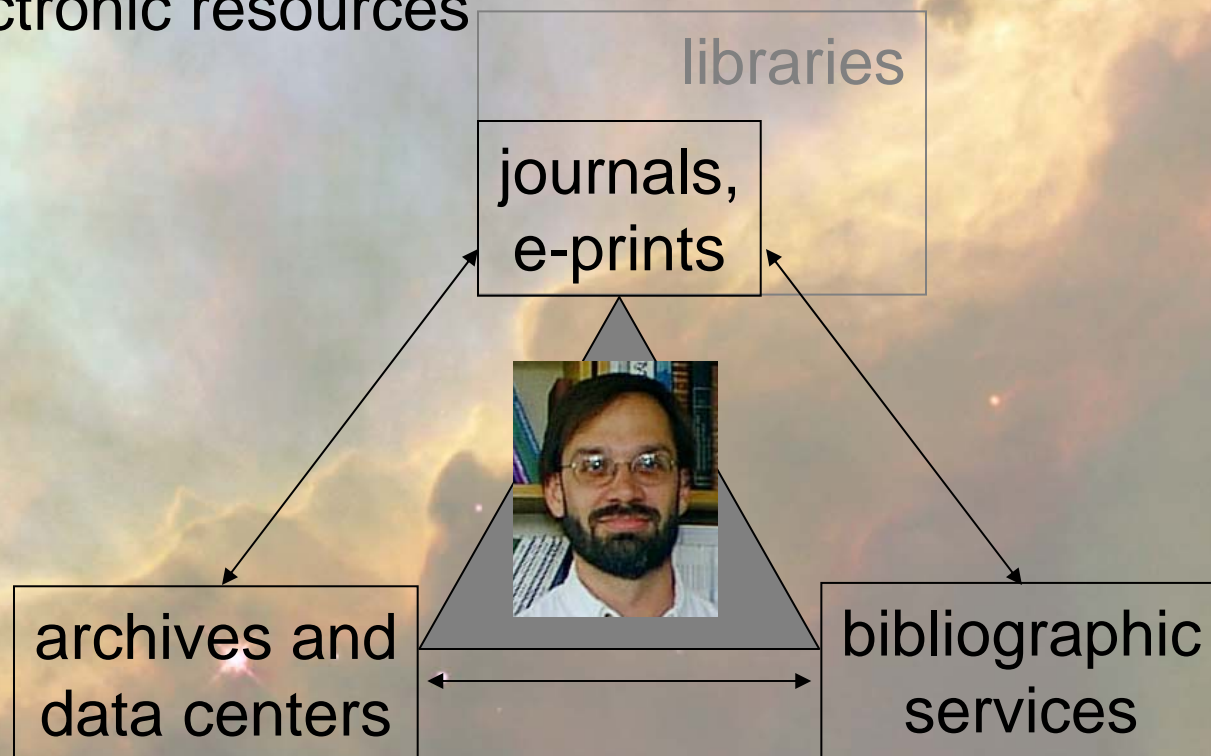
Ray Plante, National Center for Supercomputer Applications

Electronic information in astronomy

- Astronomy was one of the first scientific disciplines to pioneer e-publishing (ApJLett 1995, ApJ and AJ 1996)
- Astronomy has comprehensive e-abstract and bibliographic services
 - Astrophysics Data System, SIMBAD, NED
- Astronomy makes extensive use e-preprints on arXiv.org
- Astronomy data is archived and is generally publicly accessible
 - NASA mission archives
 - ground-based observatories (U.S., Europe, Australia, etc.)
 - data centers (catalogs, tables, value-added services)

Electronic information in astronomy

- E-journals link to underlying data, and data archives link to e-journals, through a system of persistent, unique identifiers
- Astronomers interact with a set of connected electronic resources



The Virtual Observatory

- The *Virtual Observatory* is a framework for providing access to distributed data, distributed services. The VO is about *data discovery, access, and integration*, and combining data with computational services.
- Motivation:
 - The data deluge. Needs tools to locate and sift through immense collections and to correlate data from many resources. ~500 TB of data currently available.
 - Scientific discovery opportunities exist at the intersections of diverse data sets.
- Astronomy, of course! Space science, solar physics, aeronomy, seismology, oceanography, hydrology, biology, genomics, medicine. [...]ology and [...]onomy.
- Keywords: *Metadata, interoperability*

Data/Information in the VO

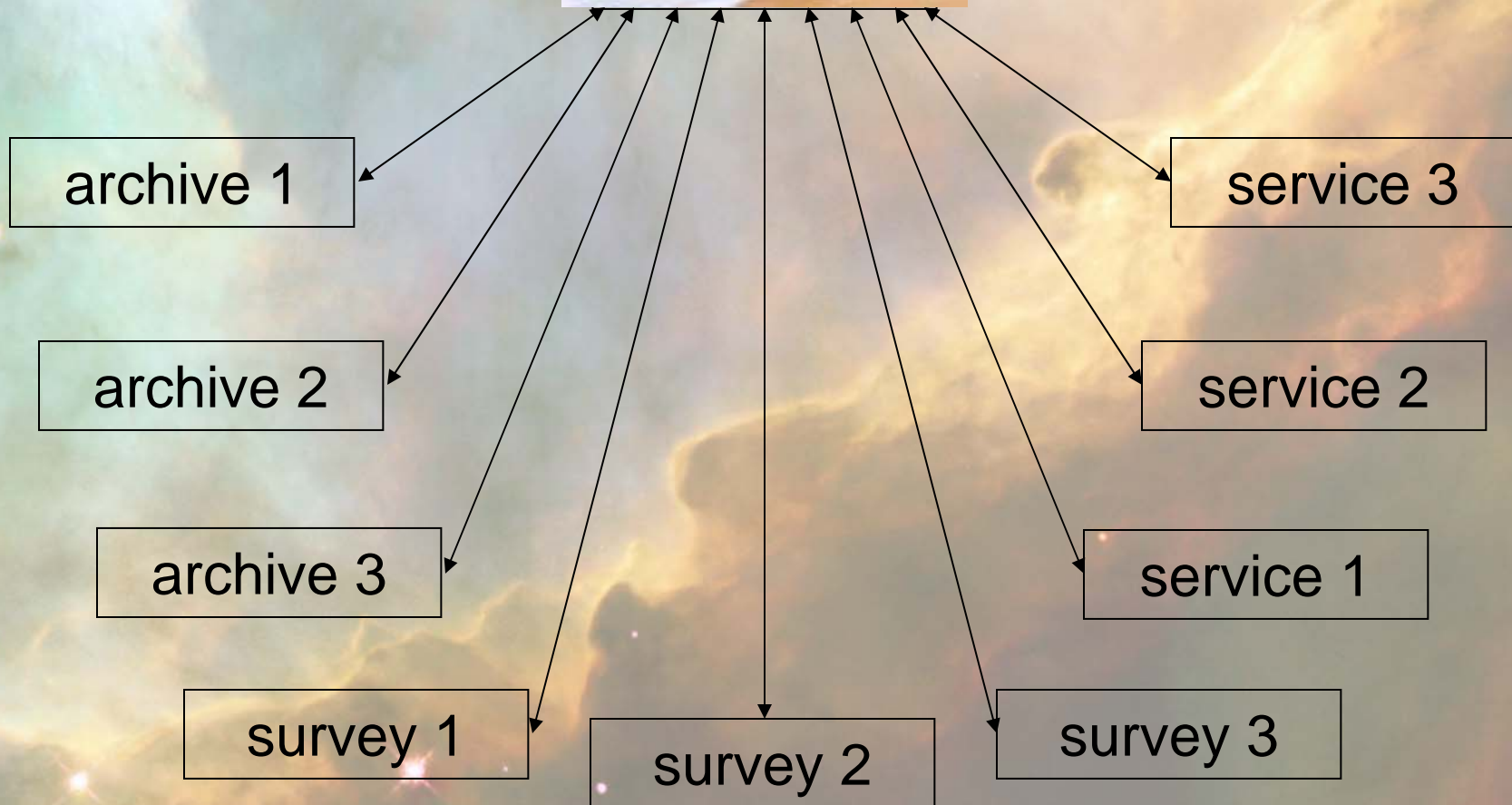
- Basic data
 - digital images, spectra, time series, catalogs, tables
- Simulations
 - models (results, computer codes, computational services)
 - virtual observations
- Analysis and interpretation
 - journals, e-preprints
 - reprocessed and enhanced data
- Name-resolution services
 - “Andromeda Galaxy”, “Messier 31”, “M31”, “NGC 224”, “UGC 454”, etc. ==> ra 00h 42m 44s, dec +41° 16' 08”
 - Geographic equivalent of “Glenn Dale, MD”, “20769”, “Prince George’s County” ==> 76° 48' 19 W, 38° 58' 36” N

*not discoverable
through text-based
search engines*

Without VO



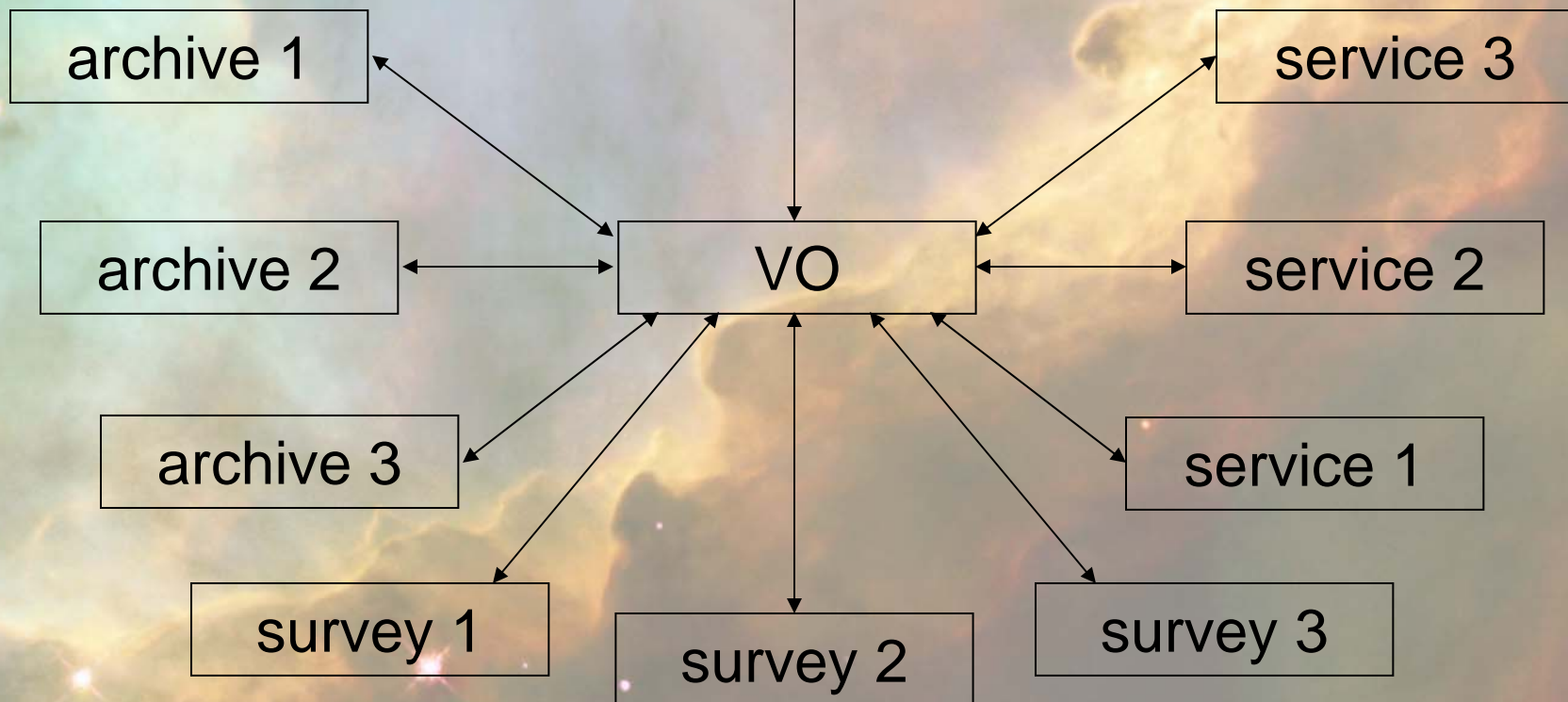
n services,
 n interfaces



With VO



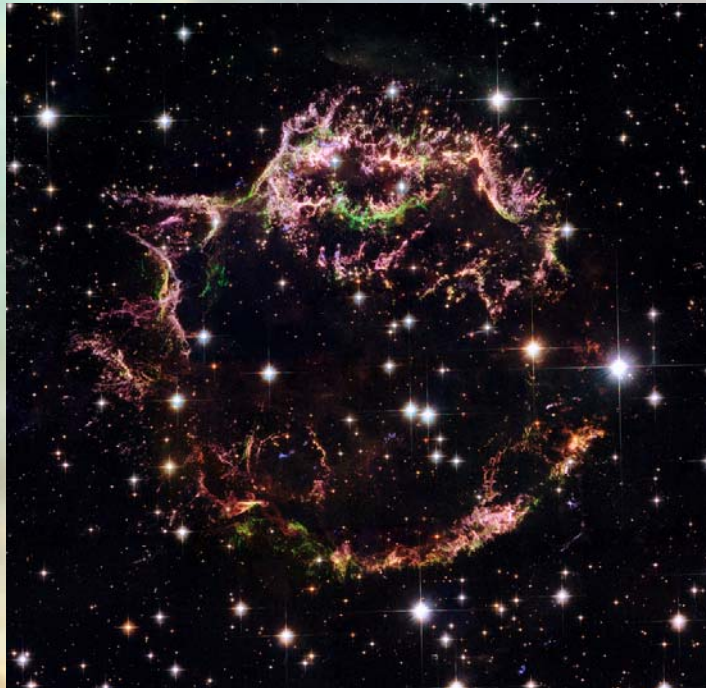
n services,
“1” interface



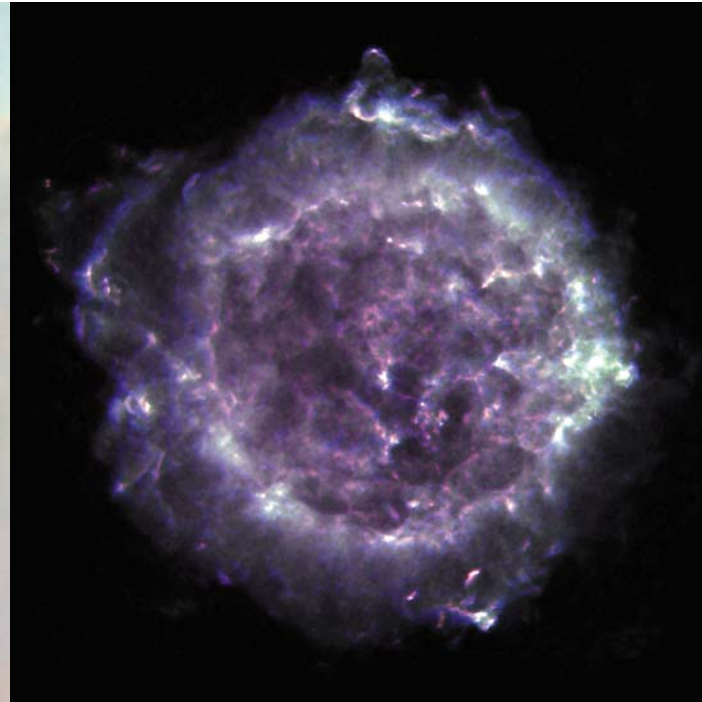
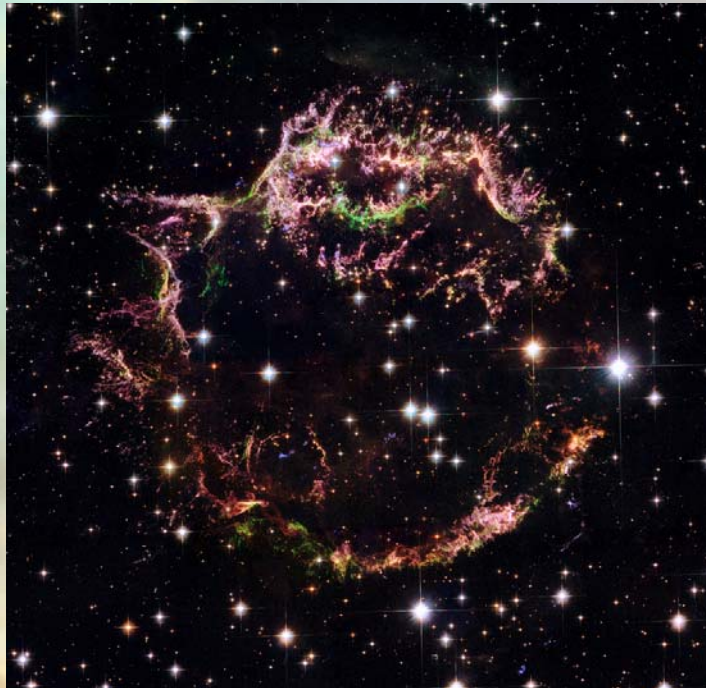
Data integration

Cas A supernova remnant

optical (HST)

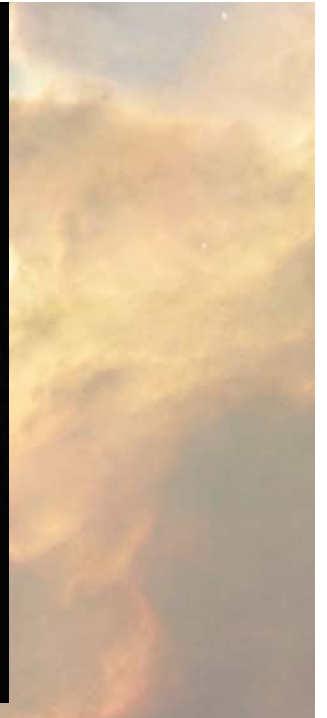
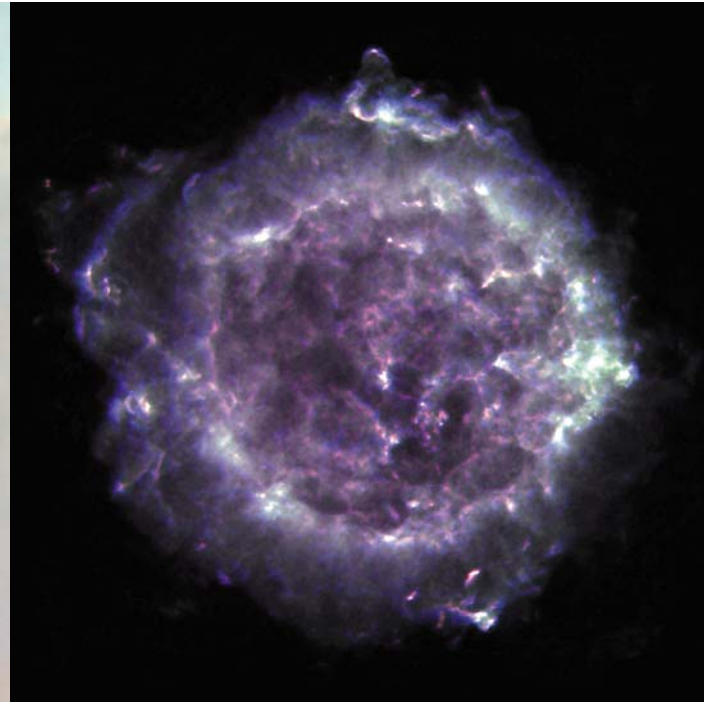
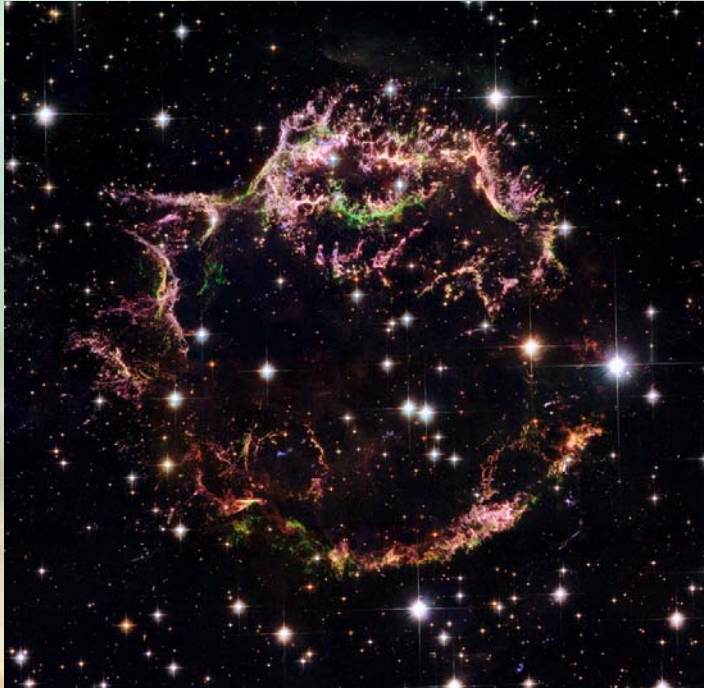


Data integration

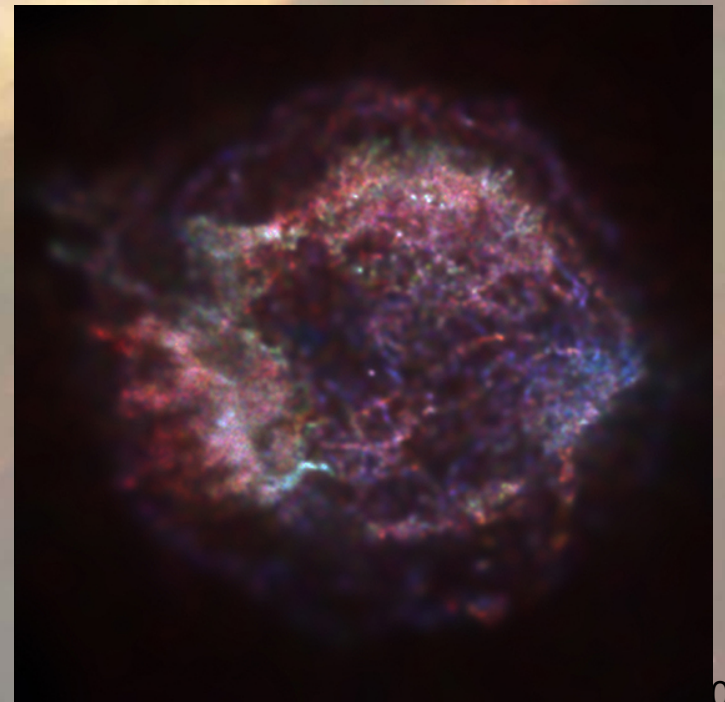


radio (VLA)

Data integration



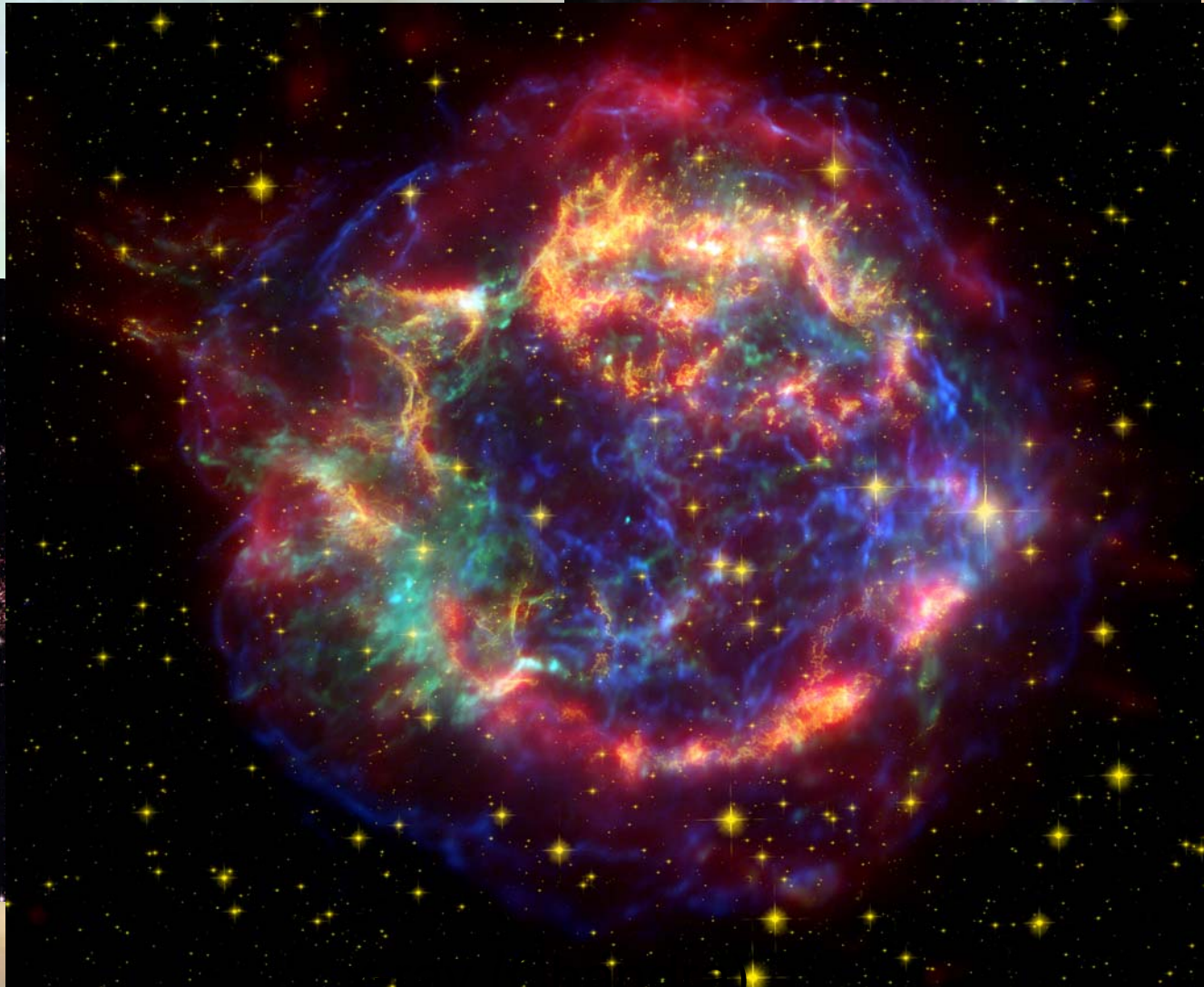
x-ray (Chandra)



10 October 2006

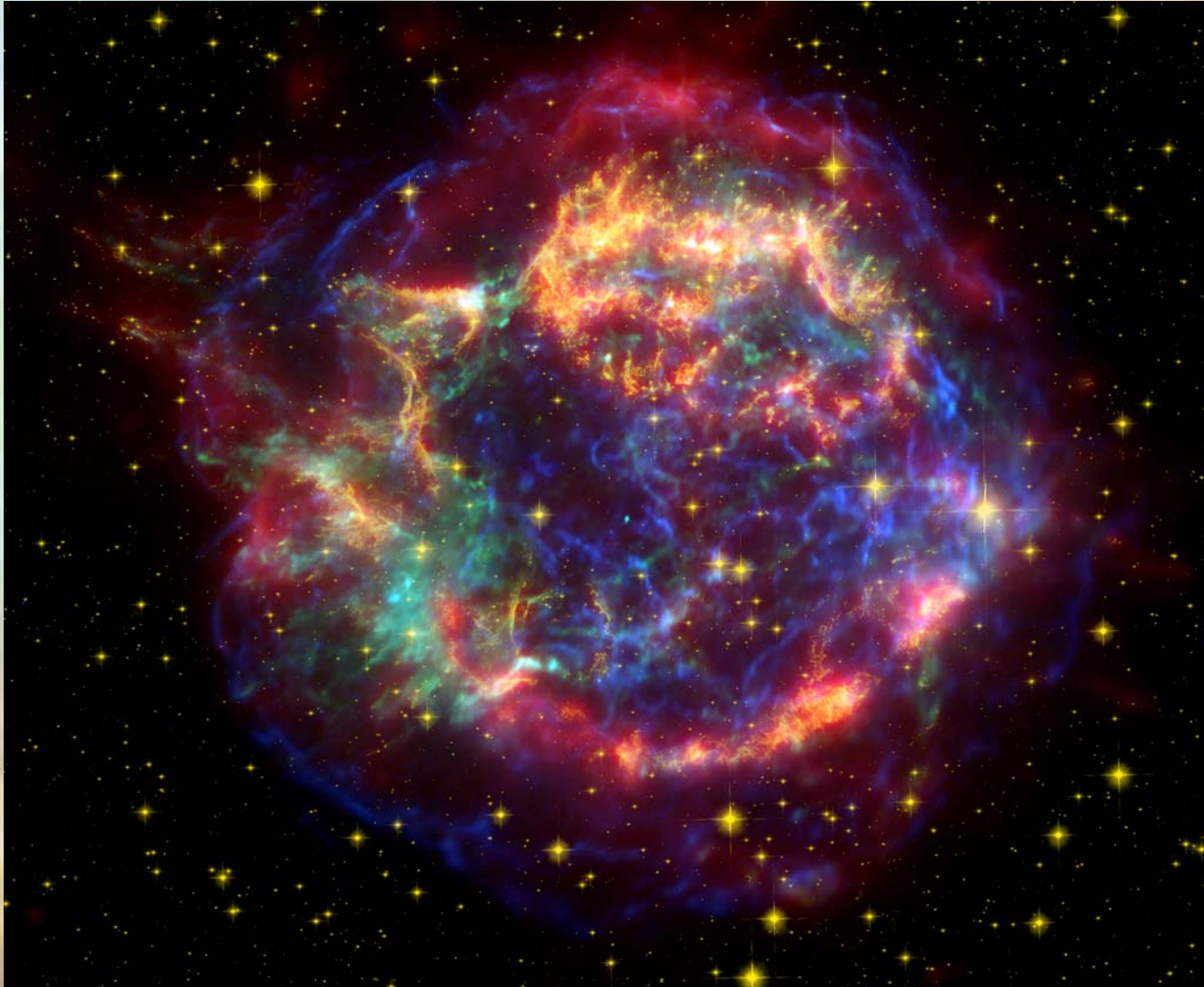
iPRES 2006

Data integration



X-ray (Chandra)

Data integration




The data preservation problem

- Research communities publish peer-reviewed journal papers that describe highly processed data.
- Long-term preservation and curation systems for digital journal content are not currently in place; *only the graphical representations of data are being saved.*
- The research cannot be verified and the results cannot be easily compared to other data in order to broaden impact.
- Public funds invested in scientific research do not have maximum return on investment. Essential legacy datasets are being lost.

Approach

- Integrate digital data management into the publication process (data capture, review, metadata tagging and validation, storage).
- Exploit emerging information technology standards for managing distributed data collections, including digital journals.
- Provide multiple access methods to digital data to maximize visibility and re-use.
- Exploit information management and curation experience in the university libraries and build on long-term institutional commitments to preservation.

Astronomy Digital Image Library



[User's Guide](#)[Step 1: Gather Files](#)

NCSA Astronomy Digital Image Library Guide: Depositing Images

Depositing Images into the Library

Contributions to the Library are made in the form of [projects](#), or collections centered around a scientific publication. A deposit contains:

- One or more **Images stored in the FITS format**.
- A **Submission Form** filled out via our Web HTML form.
- Optional number of **supplementary data files**.


Contributions can be made by following these steps:

- [Assemble FITS images and related information](#)
- [Fill out a Submission Description Form](#).
- [Send the files to the Library via Anonymous FTP](#).
- [Notify the Library of your deposit by email](#)

When we receive your email, we will load your deposit into the Library, making it available to all users. At that time, we will send back to you by email a project [codename](#) which will allow you to access your deposit without having to search the database.

Take a few minutes to look over the Project and Preview Pages to make sure your deposit was processed correctly. You can send your questions and comments to adil@ncsa.uiuc.edu.

We greatly appreciate your participation in this project!



[User's Guide](#)[Step 1: Gather Files](#)

The [Astronomy Digital Image Library](#) is a project of [Radio Astronomy Imaging Team](#) at the [National Center for Supercomputing Applications](#) on the campus of the [University of Illinois at Urbana-Champaign](#)

Contact the ADIL: adil@ncsa.uiuc.edu

ADIL query

NCSA ADIL [Get By Code](#) [Help](#)
NCSA Astronomy Digital Image Library Query Page

Click on highlighted words for [help](#) on that section of form.

[Return](#) 50 matching images starting with #1

Position: *Note: consider using the Survey Filter below with position searches*

Right Ascension:
(HH:MM:SS.SS[, HH:MM:SS.SS])

Declination:
(DD:MM:SS.SS[, DD:MM:SS.SS])

Epoch: 2000.0

Frequency:

☒ Any Frequency
☐ Search by **Waveband:**

☐ Radio ☐ Infrared ☐ Optical ☐ Ultraviolet ☐ X-ray ☐ Gamma

☐ Search by **Frequency Range:** Units:

Rest Frequency: Units:

Species: ([List of species in database](#))

Object:

Object Name: (one per line)
[List of names](#)

Object Type: (one per line)
[List of types](#)

Survey Filter: (select from menu)

☐ Use [NED object name resolution](#)
☐ Use [SIMBAD object name resolution](#)

Image Origin and Related Science:

Authors: (one per line)

Title words: (any format)

Telescopes: (one per line)
[List of telescopes](#)

ADIL query

NCSA ADIL [Get By Code](#) [Help](#)
NCSA Astronomy Digital Image Library Query Page

Click on highlighted words for [help](#) on that section of form.

[Return](#) 50 matching images starting with #1

Position: *Note: consider using the Survey Filter below with position searches*

Right Ascension: _____ **Declination:** _____
(HH:MM:SS.SS[, HH:MM:SS.SS]) (DD:MM:SS.SS[, DD:MM:SS.SS])

Epoch: 2000.0 ▾

Frequency:

☒ Any Frequency
☐ Search by **Waveband:**

☐ Radio ☐ Infrared ☐ Optical ☐ Ultraviolet ☐ X-ray ☐ Gamma

☐ Search by **Frequency Range:** _____
Units: HZ ▾

Rest Frequency: _____ Units: HZ ▾

Species: _____ ([List of species in database](#))

Object:

Object Name: (one per line) _____ **Object Type:** (one per line) _____ **Survey Filter:** (select from menu) _____

[List of names](#) [List of types](#) ▾

☐ Use [NED object name resolution](#)
☐ Use [SIMBAD object name resolution](#)

Image Origin and Related Science:

Authors: (one per line) _____ **Title words:** (any format) _____ **Telescopes:** (one per line) _____

[List of telescopes](#)

ADIL is great, but...

- Data capture and curation is separate from manuscript processing
- Data access is not integrated into the journals
- Data management is centralized

Storyboard

The Astrophysical Journal, 644:759-768, 2006 June 20
© 2006. The American Astronomical Society. All rights reserved. Printed in U.S.A.

Evolution of the Color-Magnitude Relation in High-Redshift Clusters: Early-Type Galaxies in the Lynx Supercluster at $z \sim 1.26$

Simona Mei,¹ Brad P. Holden,² John P. Blakeslee,^{1,3} Piero Rosati,⁴ Marc Postman,^{1,5}
Myungkook J. Jee,¹ Alessandro Rettura,^{4,6} Marco Sirianni,⁵ Ricardo Demarco,¹ Holland C. Ford,¹
Marijn Franx,⁷ Nicole Homeier,¹ and Garth D. Illingworth²

Received 2005 October 10; accepted 2006 February 24

ABSTRACT

Color-magnitude relations (CMRs) have been derived in two high-redshift clusters, RX J0849+4452 and RX J0848+4453 (with redshifts of $z = 1.26$ and 1.27 , respectively), that lie in the highest redshift cluster superstructure known today, the Lynx Supercluster. The CMR was determined from ACS imaging in the WFC F775W (λ_{775}) and F850LP (λ_{850}) filters combined with ground-based spectroscopy. Early-type cluster candidates have been identified according to the Postman et al.

Stanford et al. (2001). Recently, deep, panoramic multicolor (λ_{775} and λ_{850} bands) imaging around these two central clusters identified seven galaxy groups (Nakata et al. 2005) with photometric redshift $z_{\text{phot}} \sim 1.26$. This makes the Lynx region a unique laboratory, being the only supercluster observed at such a high redshift today, and for this reason, one of the best regions at $z > 1$ in which we can study properties of evolving galaxies within a structure that is still assembling, and in different environments.

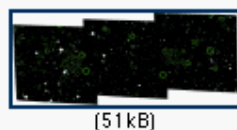


Fig. 1 *Chandra* X-ray contours overlaid on the ACS color composite image for Lynx E (on the left) and Lynx W (on the right). The contours are adaptively smoothed with a minimum significance of 3σ . We refined the alignment of the *Chandra* image with respect to the ACS using the X-ray point sources.

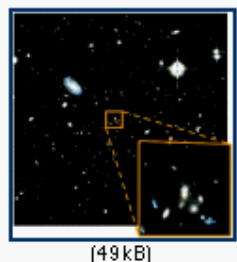
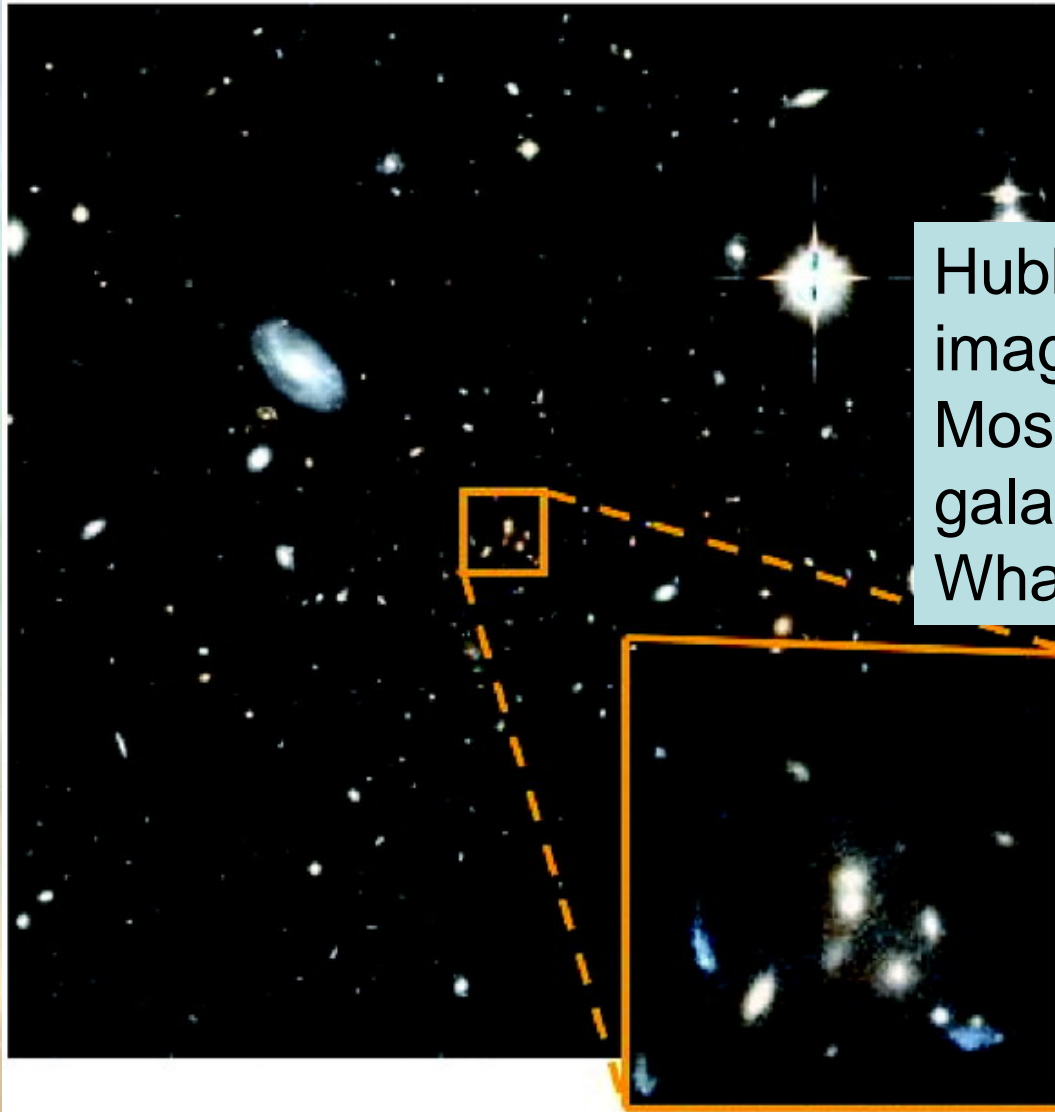


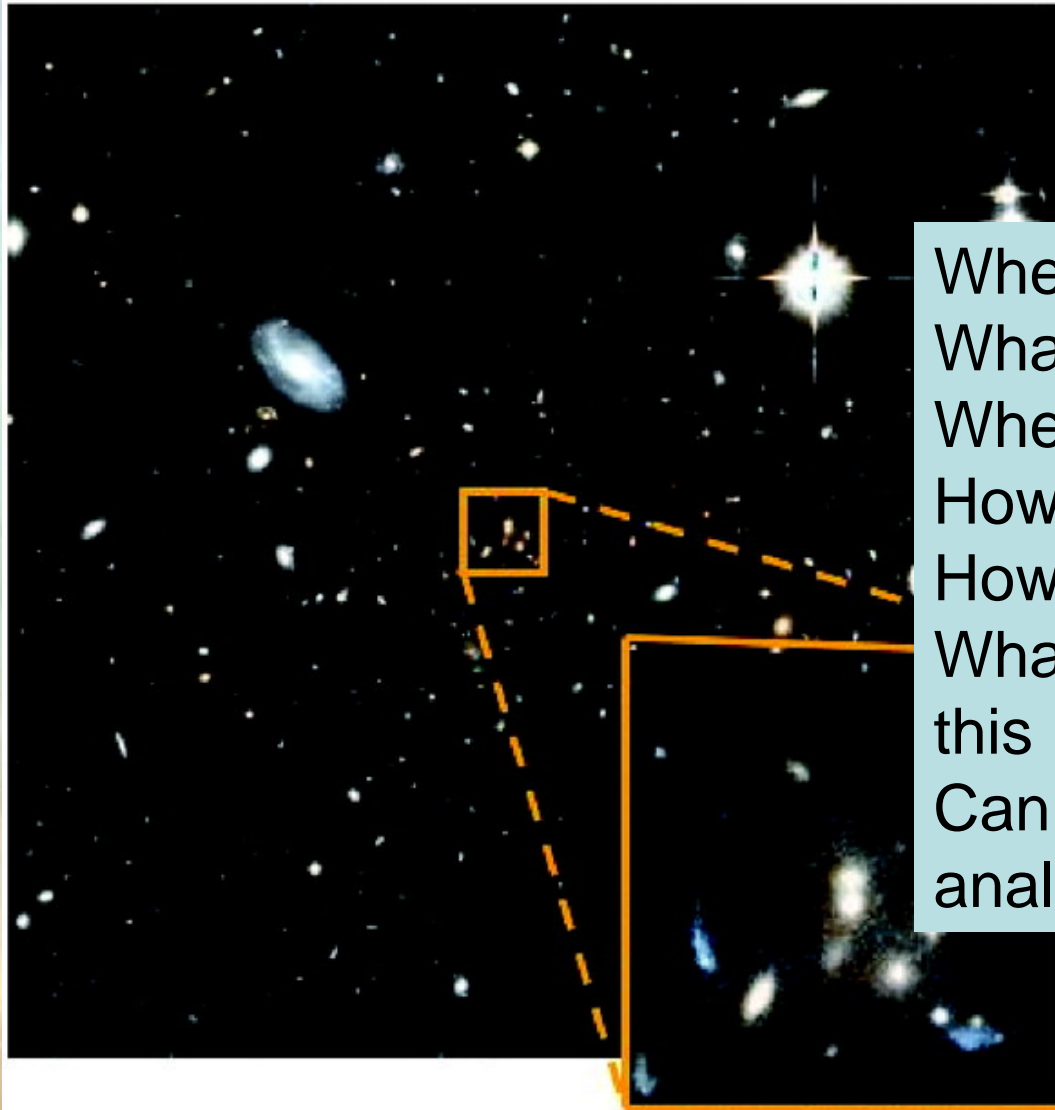
Fig. 2 Lynx E ACS image (scale is $1' \times 1'$). The central ongoing merger is magnified to also show a gravitational arc and its likely counterimage.

Storyboard



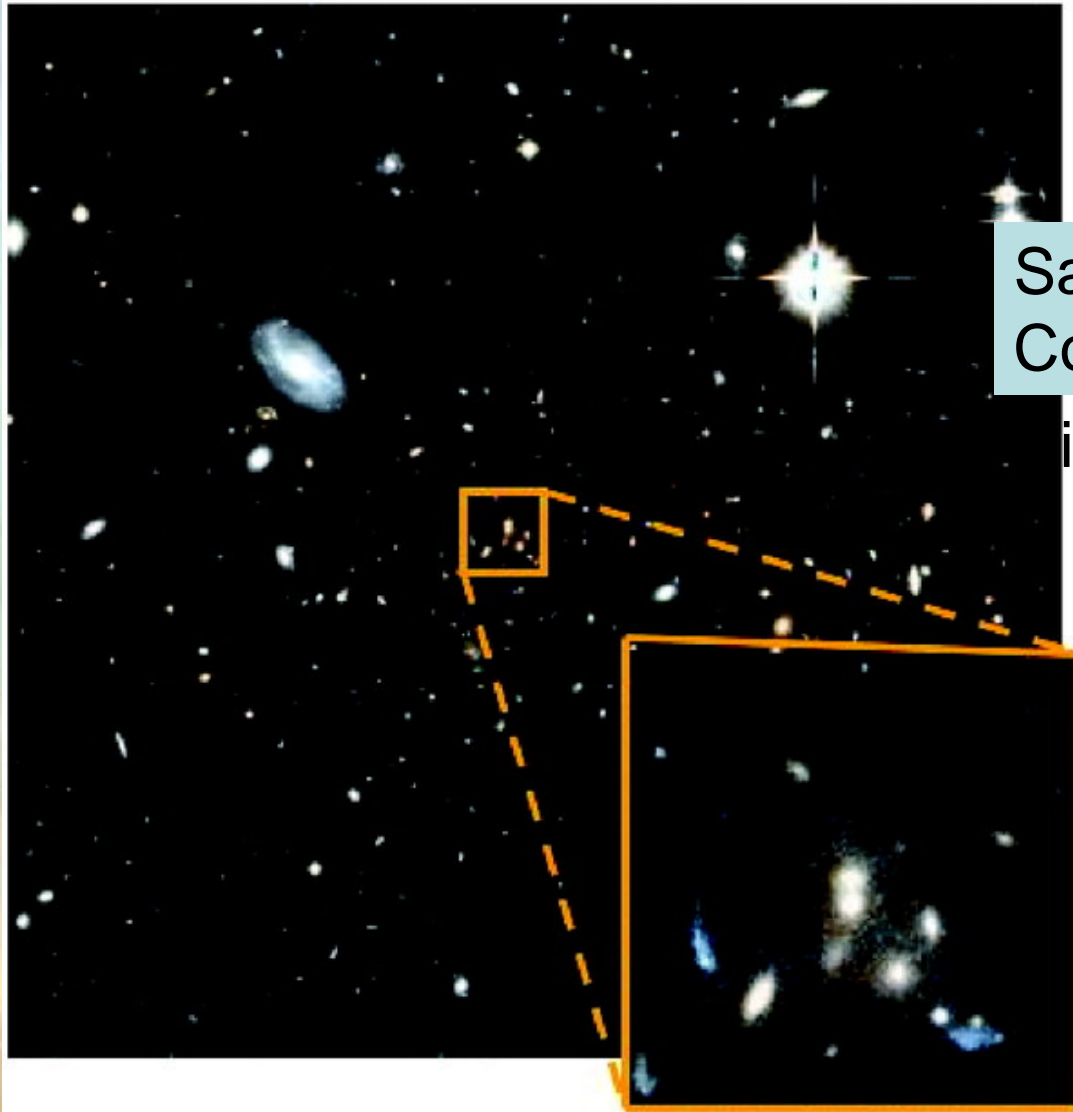
Hubble Space Telescope
image.
Most distant cluster of
galaxies known.
What more can I find out?

Storyboard

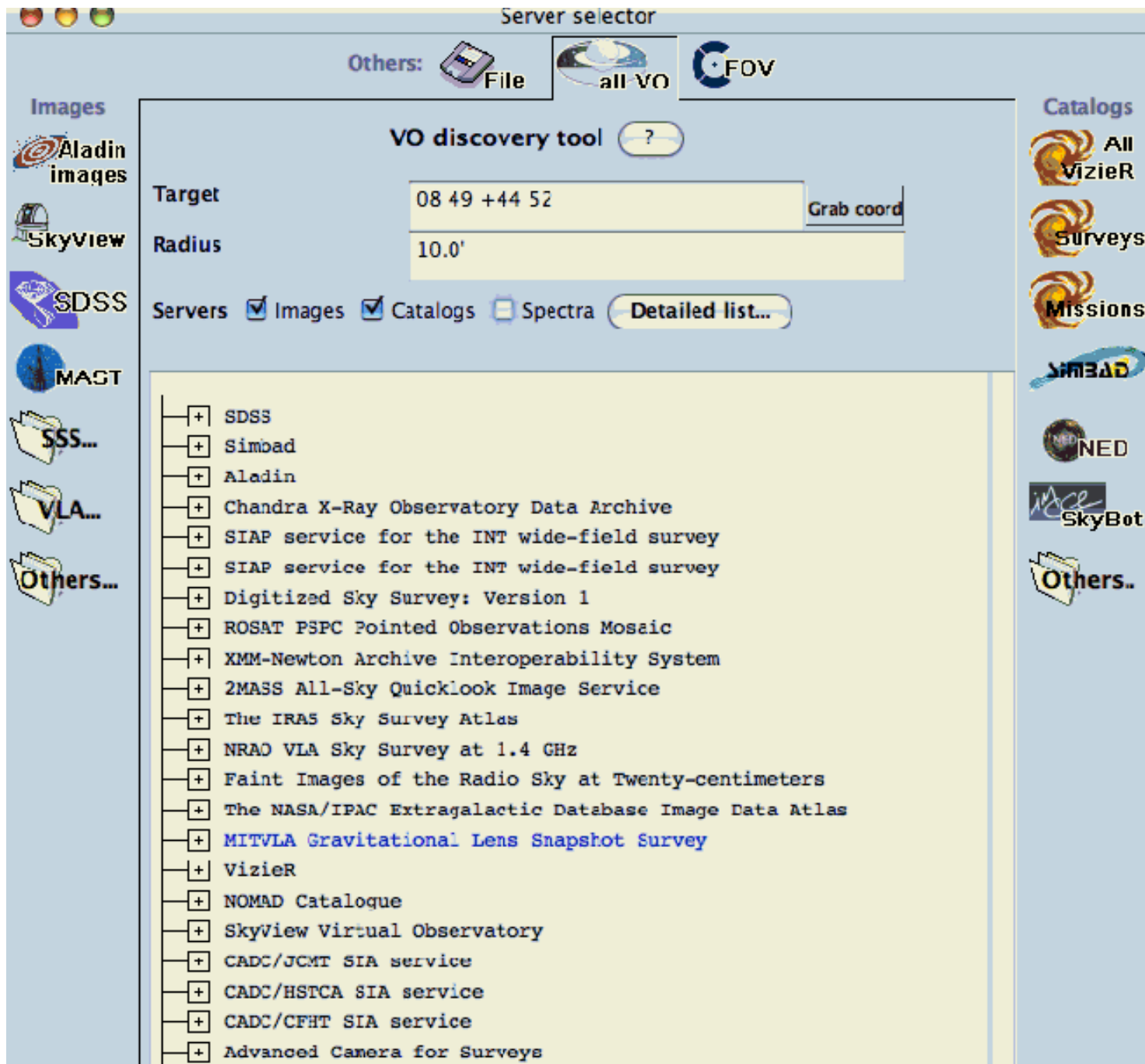


Where is this?
What is the image scale?
Where is north?
How bright is the star?
How bright is the galaxy?
What else is known about this region?
Can I trust the data analysis in this paper?

Storyboard



Save file
Copy to my VOSpace
display and compare



Aladin V3.6 multiview

Load... Save... Tools... Print... Help... Quit

Position J2000 Pixel 8 bits 037 / 255

.F850LPJ8PV022TV_DRZ

Images

Aladin images

SkyView

SDSS

MAST

SSS...

VLA...

Others...

Target

Radius

Servers ☒ Images

- + SDSS
- + Simbad
- + Aladin
- + Chandra X-Ray
- + SIAP service
- + SIAP service
- + Digitized Sky
- + ROSAT PSPC
- + XMM-Newton
- + 2MASS All-Sky
- + The IRAS Sky
- + NRAO VLA Sky
- + Faint Images
- + The NASA/IPAC
- + MITVLA Gravitational
- + VizieR
- + NOMAD Catalog
- + SkyView Virtual
- + CADC/JCMT Sky
- + CADC/HSTCA Sky
- + CADC/CFHT Sky
- + Advanced Camera

15"

1.67' x 1.67'

N

E

multiview

select

dist

draw

tag

text

filter

rgb

blink

rsamp

cont

zoom

mgls

hist

prop

del

J.AJ.123.222

J.ApJ.S.142.1

CL 0848.6+

ROSAT PSPC

F606W.U6F

F606W.U6F

F606W.U6F

F606W.U6F

F606W.U6F

F850LPJ8P

Sloan Digital

4.74' x 4.62'

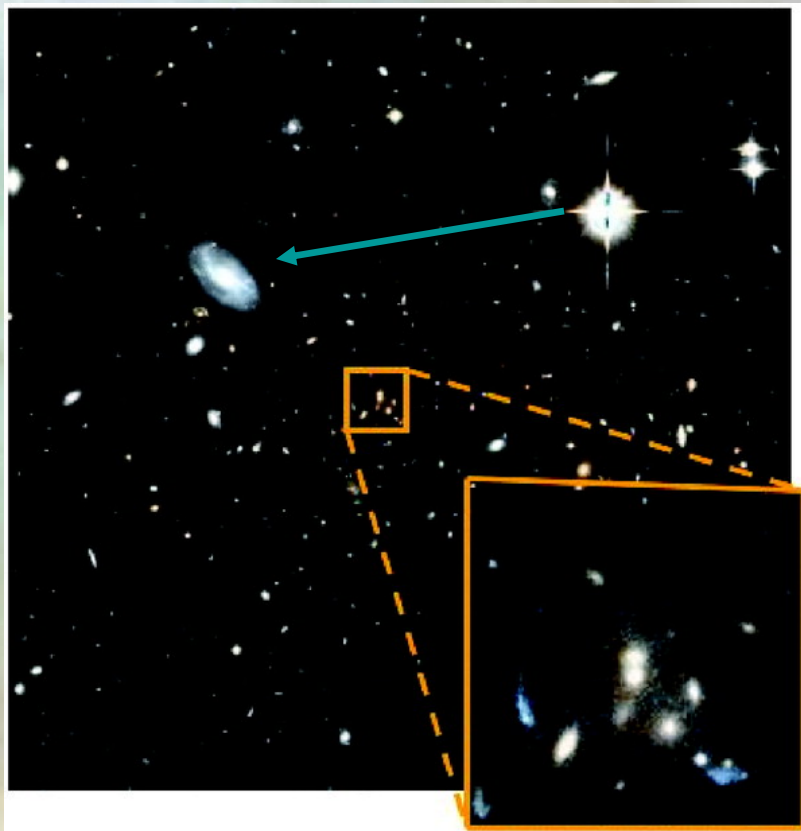
Zoom 1/4x

10 October 2006

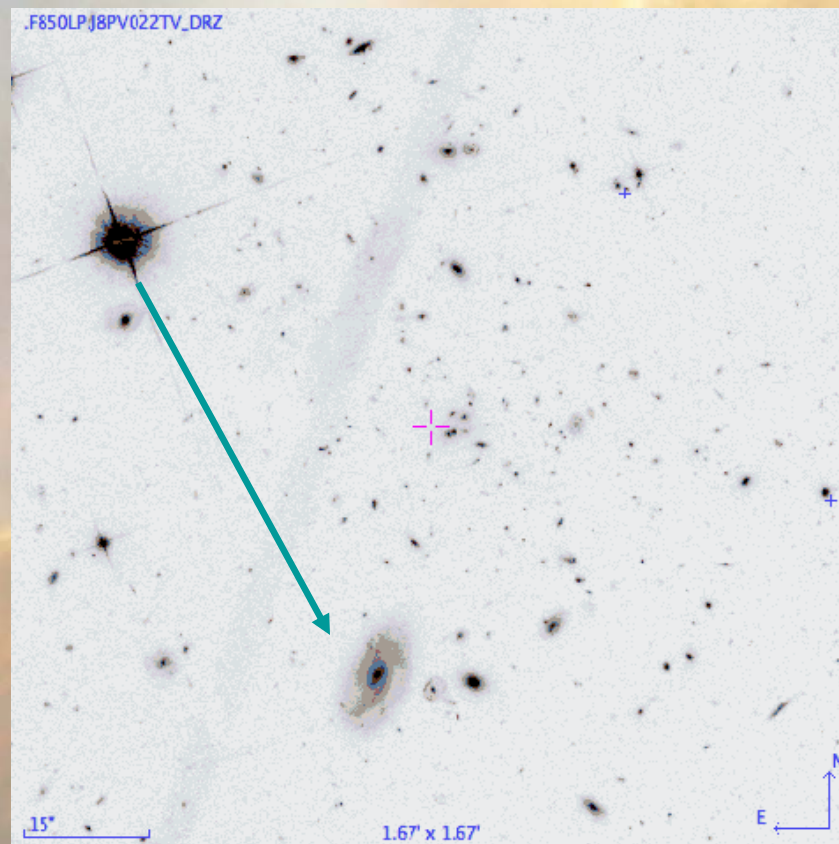
(c)1999–2006 ULP/CNRS – Centre de Données astronomiques de Strasbourg

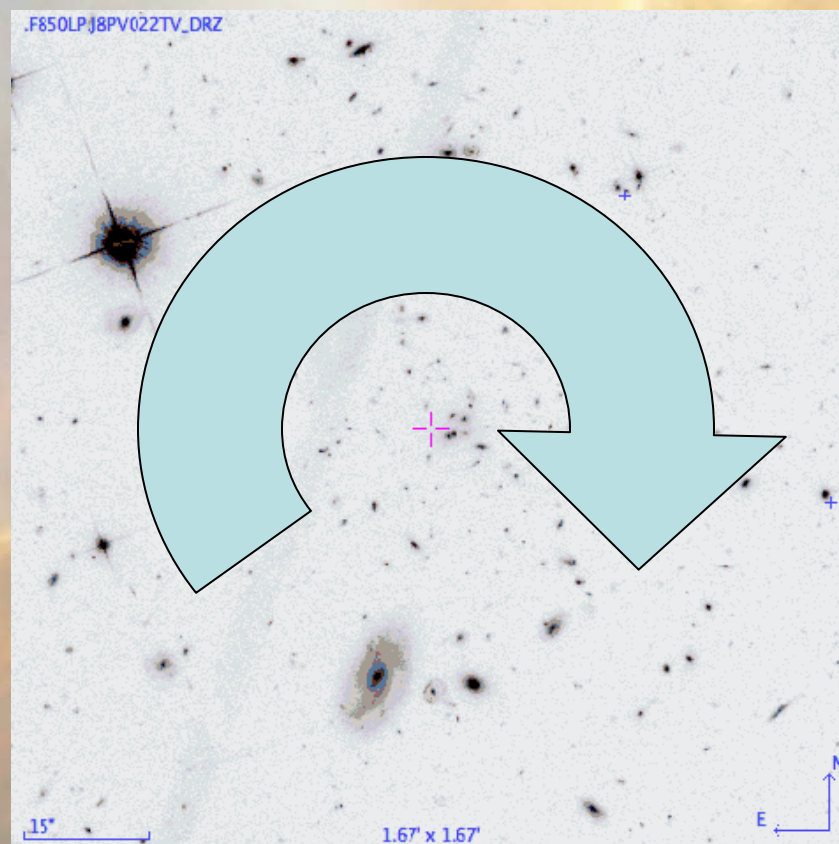
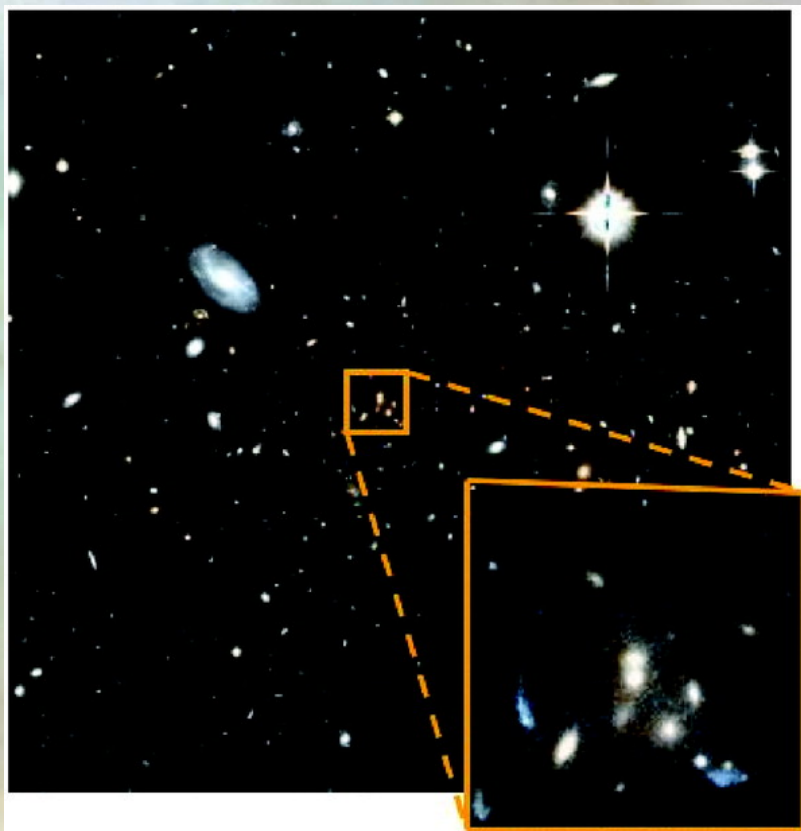
10 planes, 1 view, 266Mb

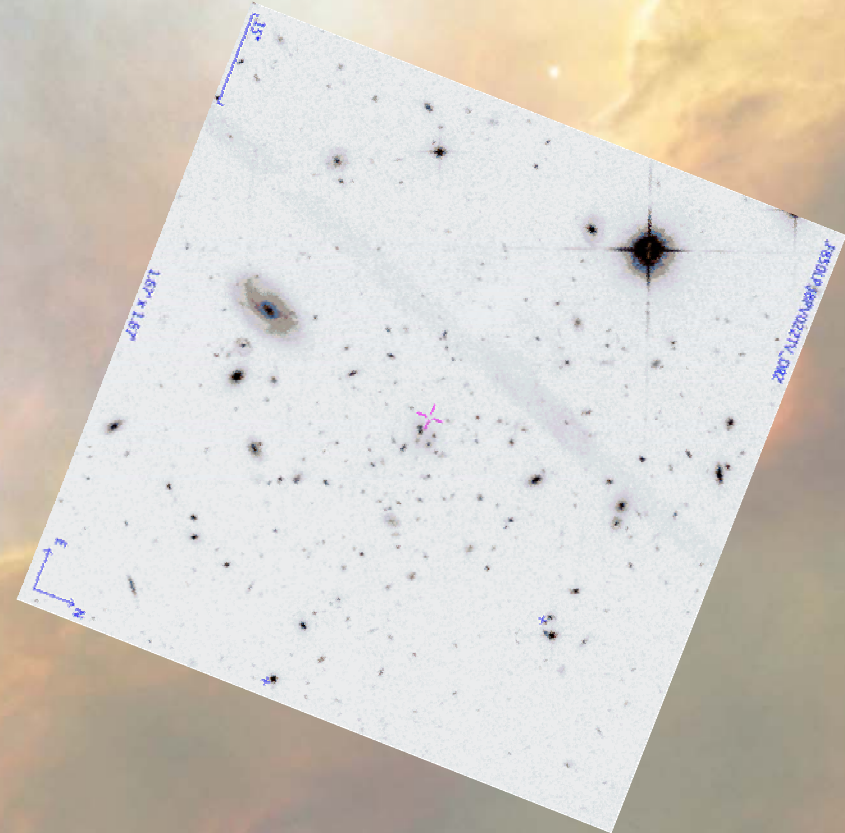
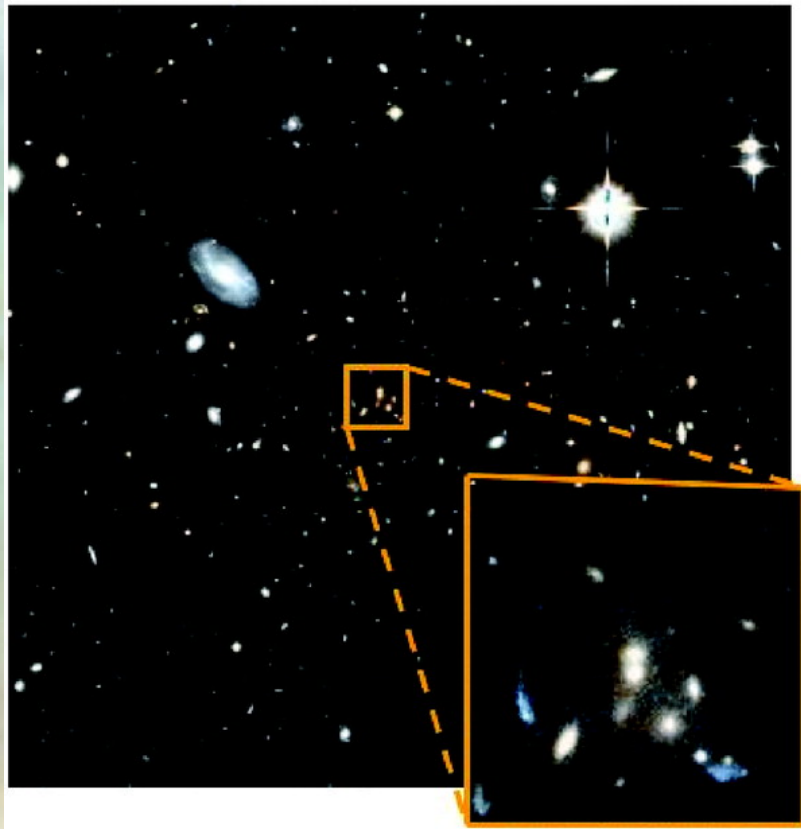
Journal...



Archive...





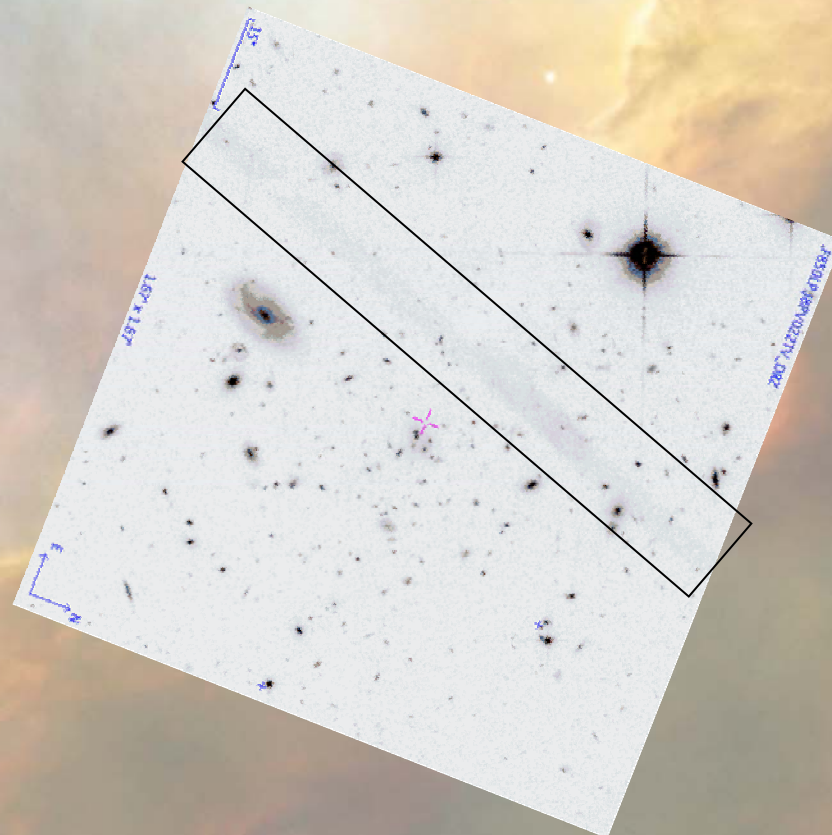
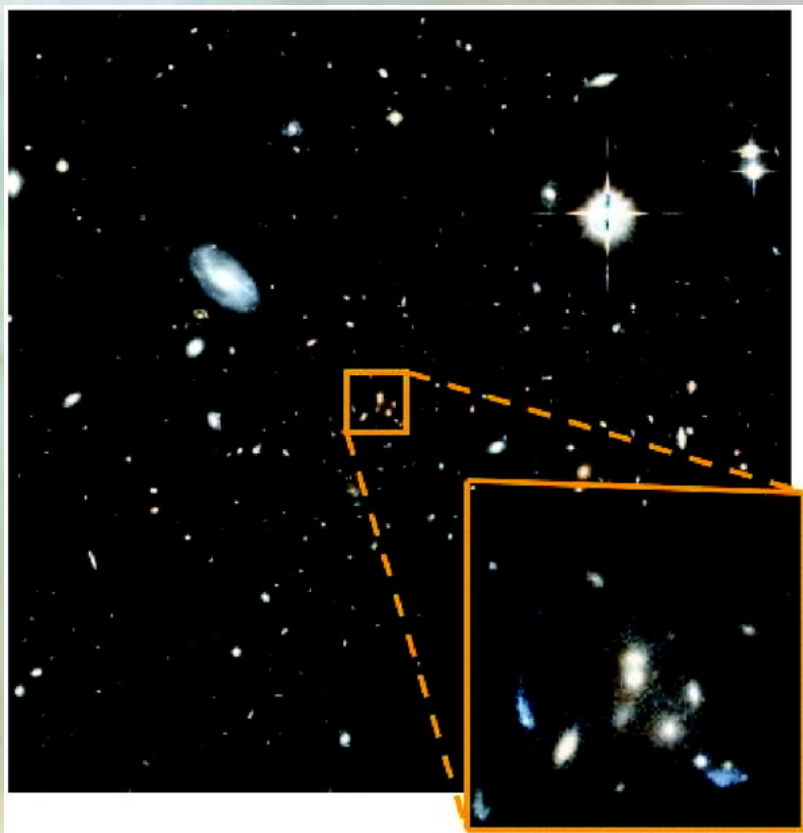




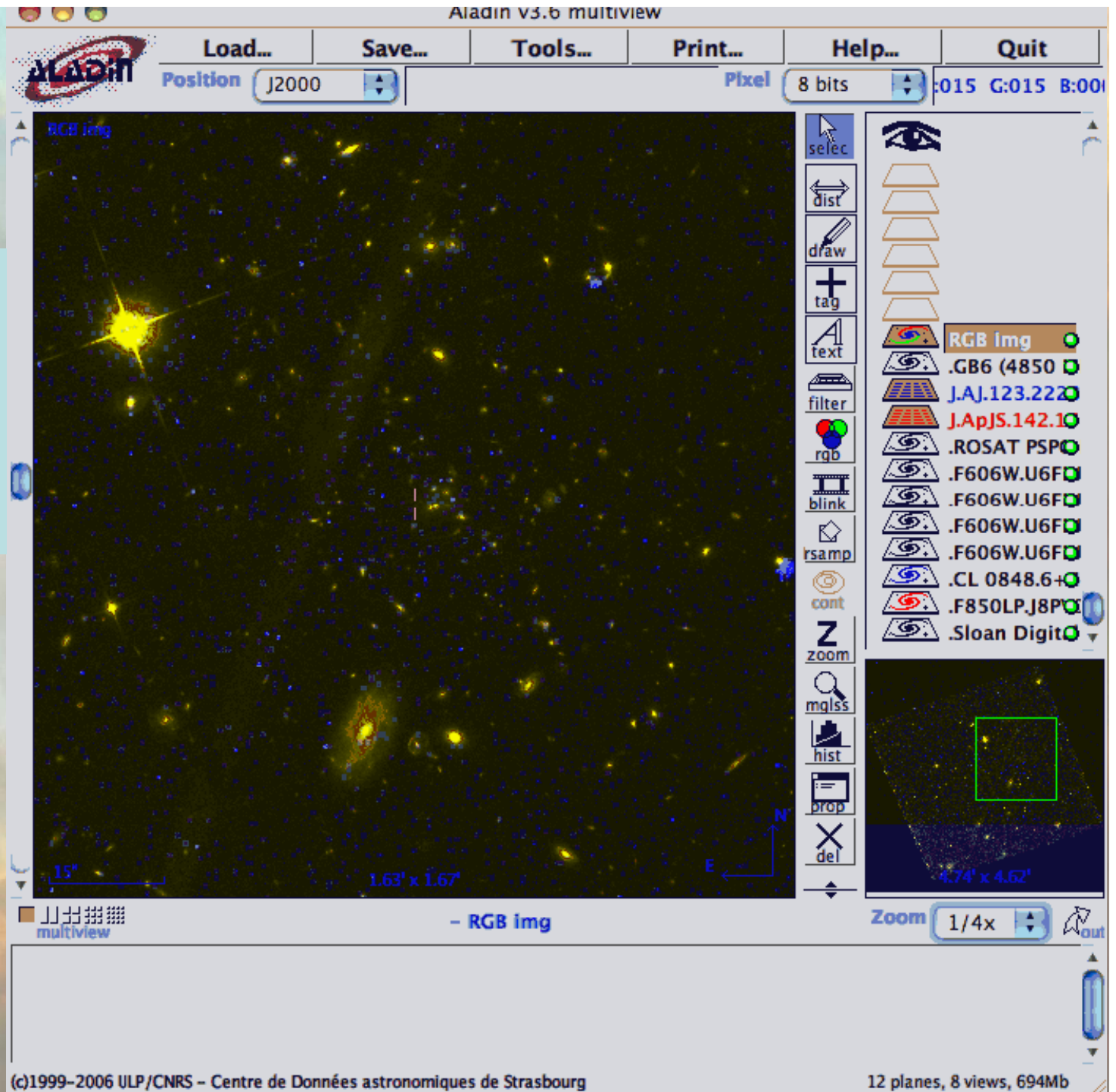
10 October 2006

iPRES 2006

27

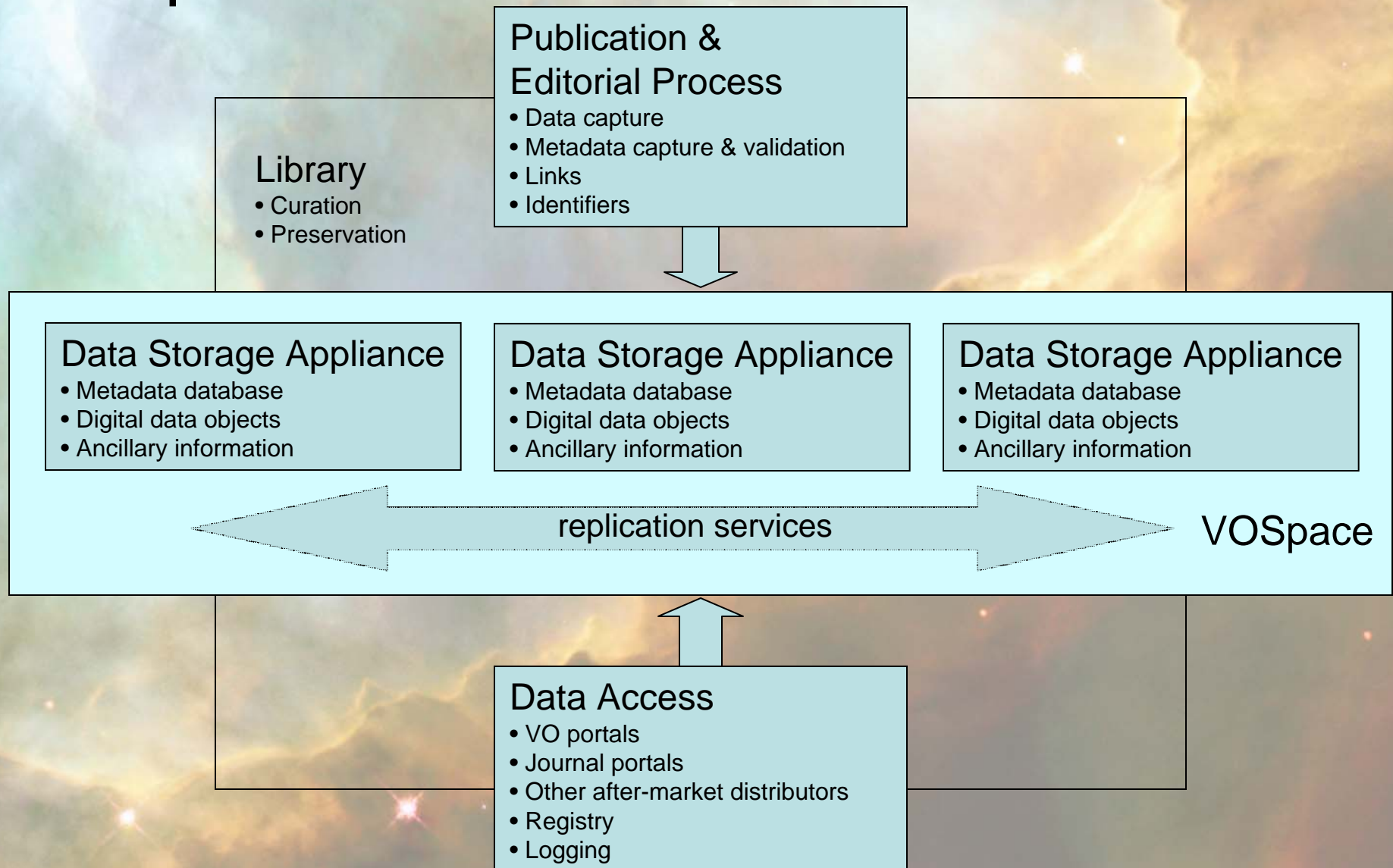


Is there any
X-ray
emission from
this cluster of
galaxies?



10 October 2006

Components



A prototype project

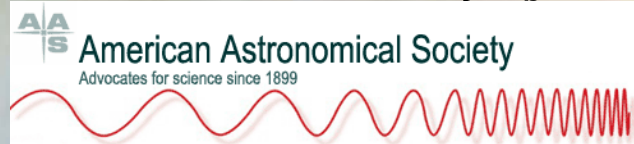
- Implement end-to-end prototype using astronomy scholarly publications as a test-bed.
- Understand operational costs and develop long-term business plan for preservation of peer-reviewed journal content and associated supporting data.
- Develop associated policies affecting data accessibility (e.g., move toward requiring digital data availability as requirement for publication).
- Utilize commodity open-source technologies and partner with Virtual Observatory to maximize return on investment, flexibility, adaptability.
- Long-term: evaluate impact on citations and productivity resulting from having ready access to digital data.

A prototype project

- Tasks (partners)
 - Metadata definition (VO, library)
 - Content management tool evaluation/selection (Fedora) (VO, library)
 - Physical storage and replication (VO, library, publisher)
 - Publication process revisions and testing (publisher, editorial staff)
 - Policy development (editorial staff, professional society)
 - Business model development (publisher, professional society)

Current collaborators

- American Astronomical Society (journals, editors)



- The University of Chicago Press (publisher for the AAS journals)



- The Johns Hopkins University-Sheridan Library and Cornell University Library (information management, curation & e-pu



- The National Virtual Observatory project (representatives from JHU, Space Telescope Science Institute, and the National Center for Supercomputing Applications)



Status

- Support committed or promised from
 - UK JISC (Joint Information Systems Committee)
 - SPARC (Scholarly Publishing and Academic Resources Coalition)
 - Microsoft
 - TeraGrid
 - NVO
 - IMLS
- Begin development in fall of 2006

Digital data discovery and access is essential for the research community

- Data re-use, with provenance
- Optimization of public investment in science
- Increasing the discovery space
- Creation of a research legacy
- Integrity in scientific publication