New Astronomy in a Virtual Observatory

S. G. Djorgovski (Caltech)

Presentation at the NSF Symposium on Knowledge Environments for Science, Arlington, 26 Nov 02

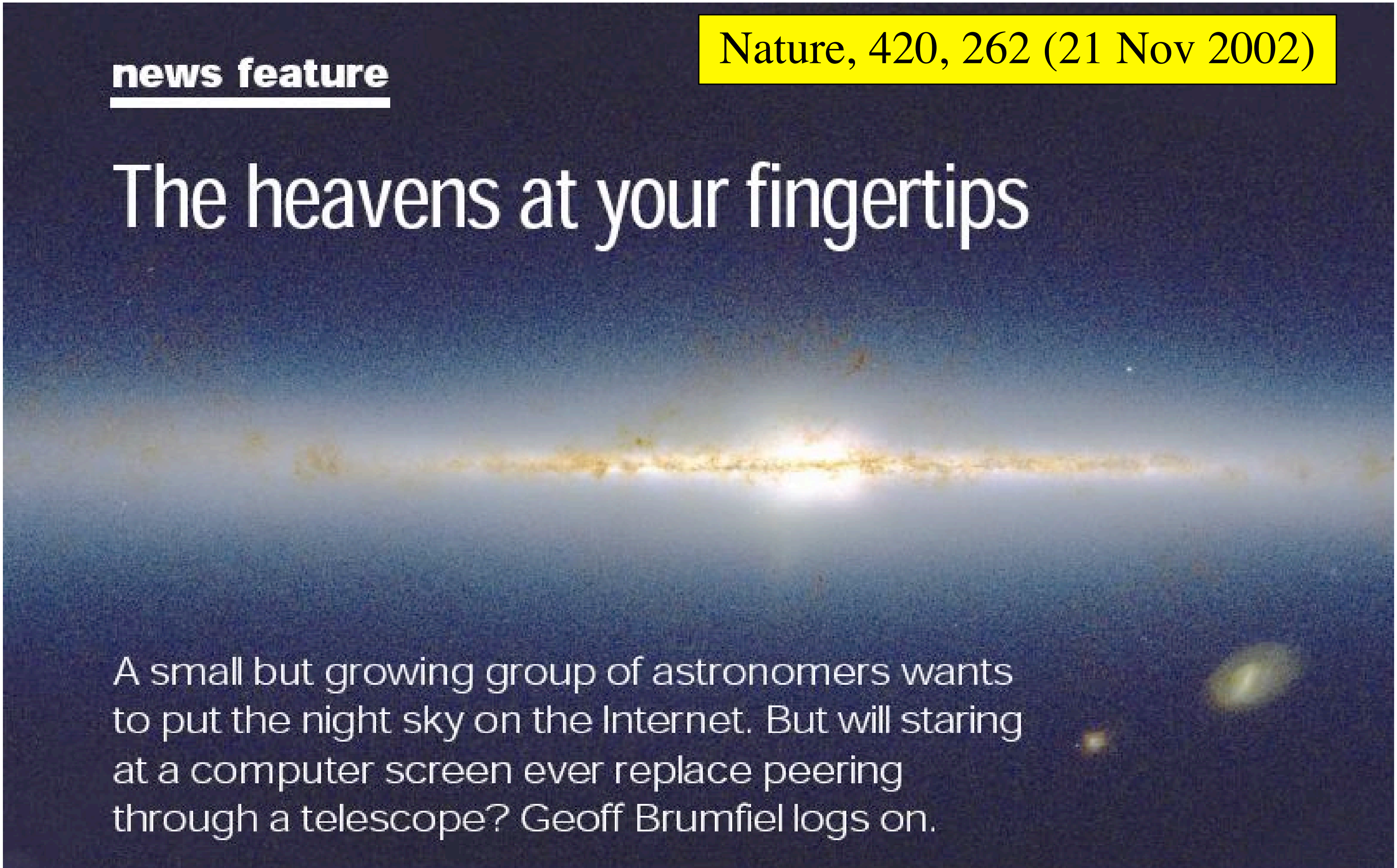
- The concept of a Virtual Observatory (VO)
- Technological opportunities, scientific needs
- A new type of a scientific organization / environment
- Towards a qualitatively different science in the era of information abundance

For more details and links, please see
http://www.astro.caltech.edu/~george/vo/

Nature, 420, 262 (21 Nov 2002)

The heavens at your fingertips

A small but growing group of astronomers wants to put the night sky on the Internet. But will staring at a computer screen ever replace peering through a telescope? Geoff Brumfiel logs on.



Pune in India is no place for an observatory. Located 160 kilometres south-east of Mumbai, formerly Bombay, the city is pounded by monsoon rains for four months of every year. Besides heavy cloud

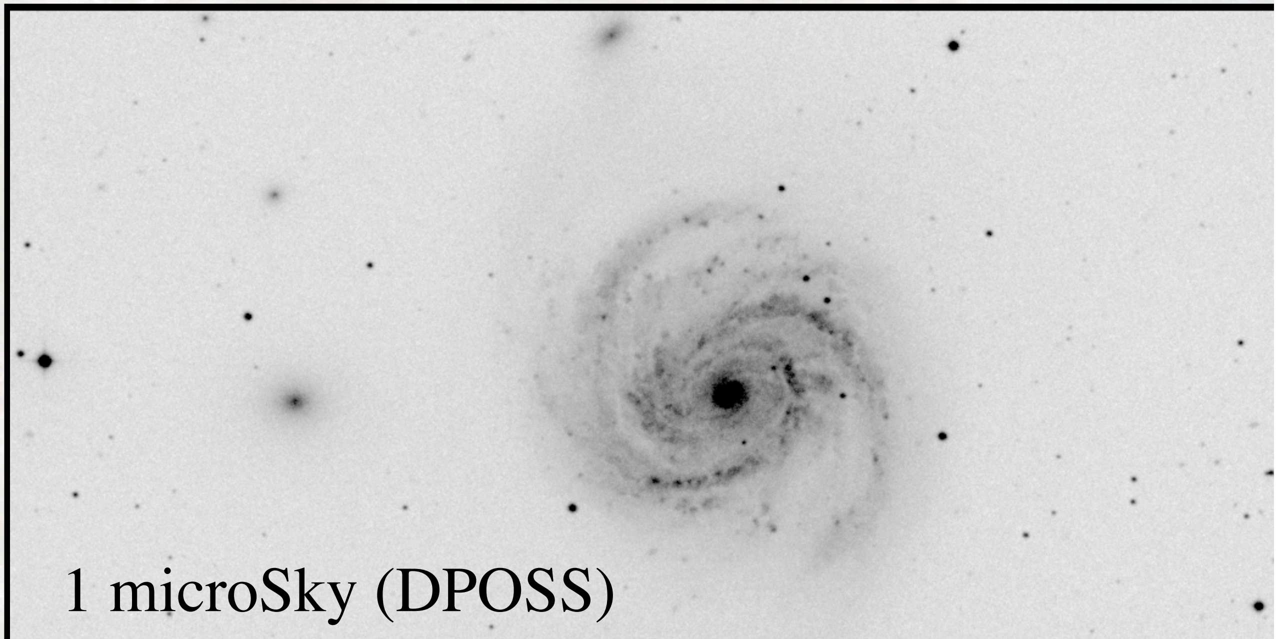
archiving information on objects outside our Galaxy in 1987, and is now linked to data from the Hubble Space Telescope.

In recent years, the amount of online data has been further boosted by a new generation

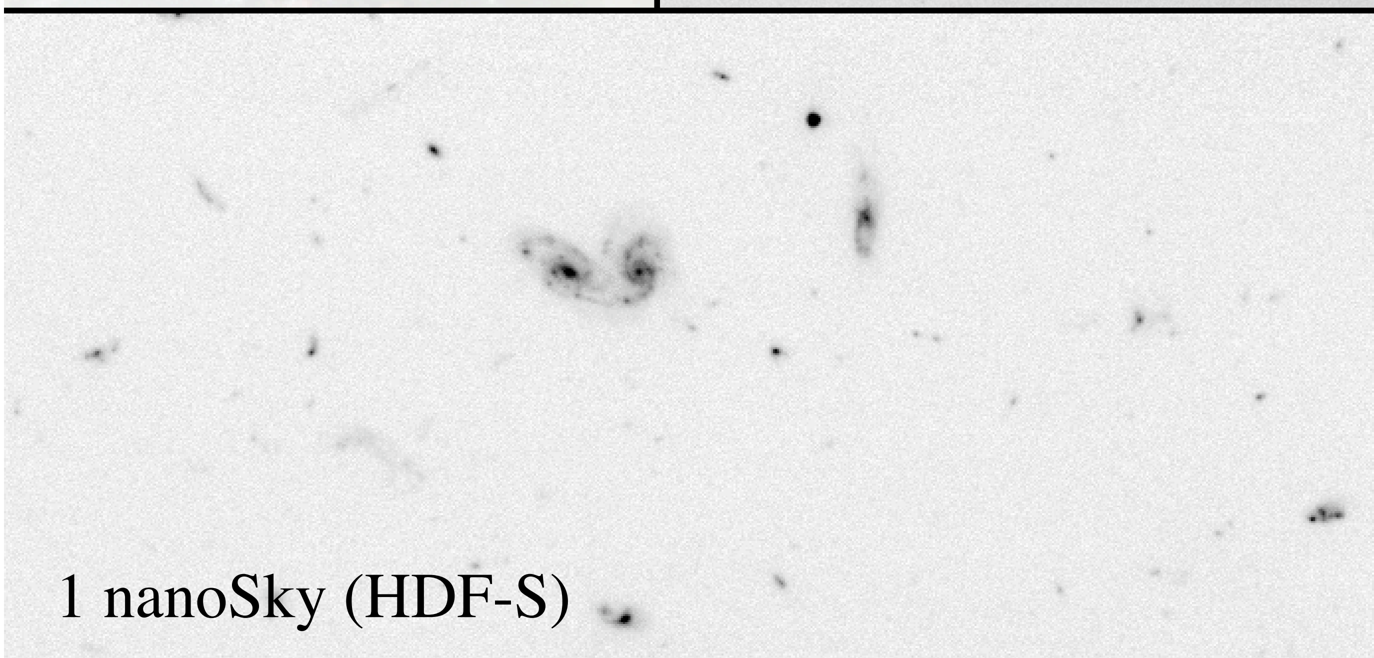
are being developed — Europe's As
Virtual Observatory, headed by Qu
National Virtual Observatory, ar
AstroGrid — with several smaller p
taking shape (see table opposite).

Astronomy is Facing a Major Data Avalanche:

Multi-Terabyte (soon: multi-PB) sky surveys and archives over a broad range of wavelengths …

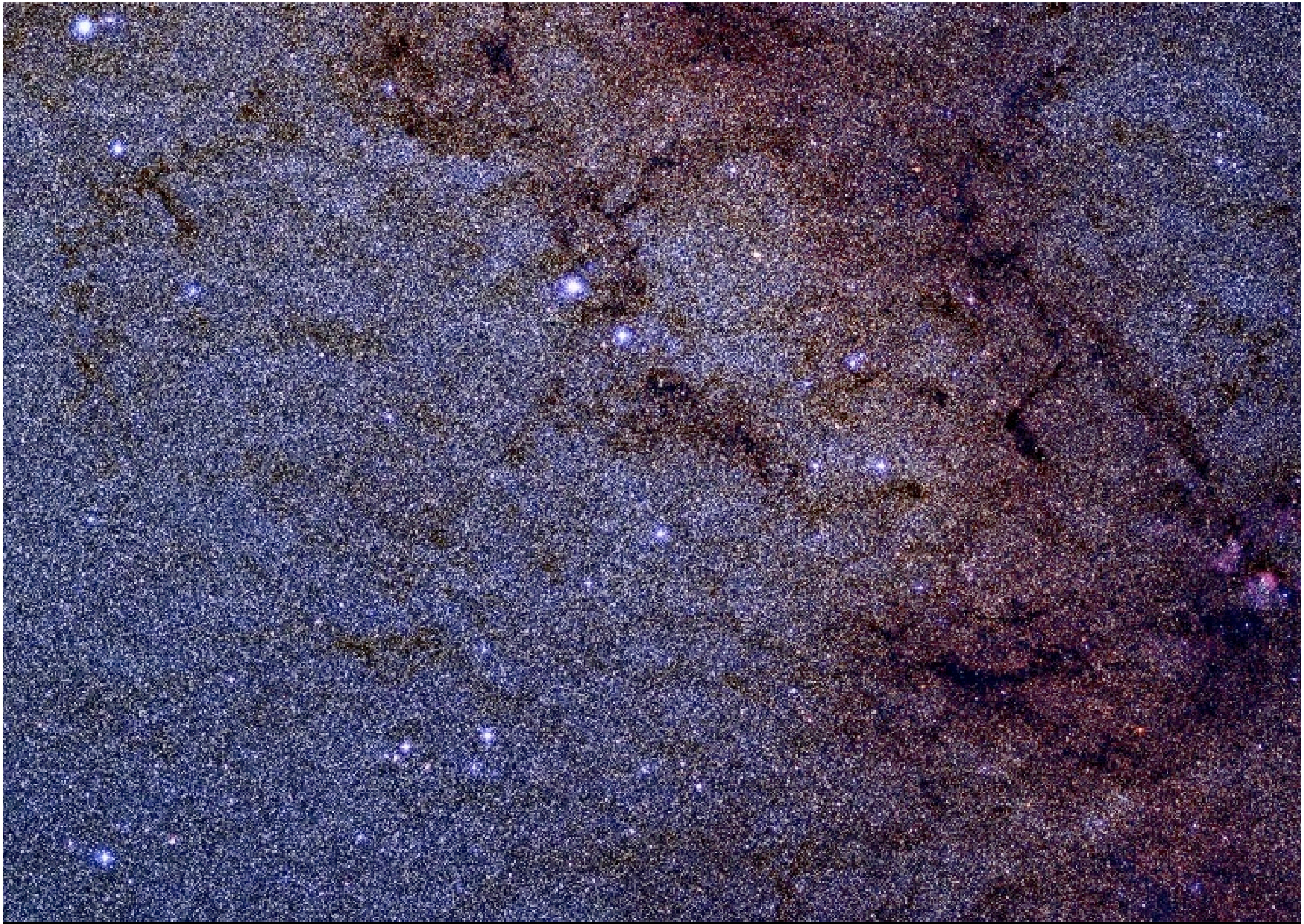


1 microSky (DPOSS)



1 nanoSky (HDF-S)

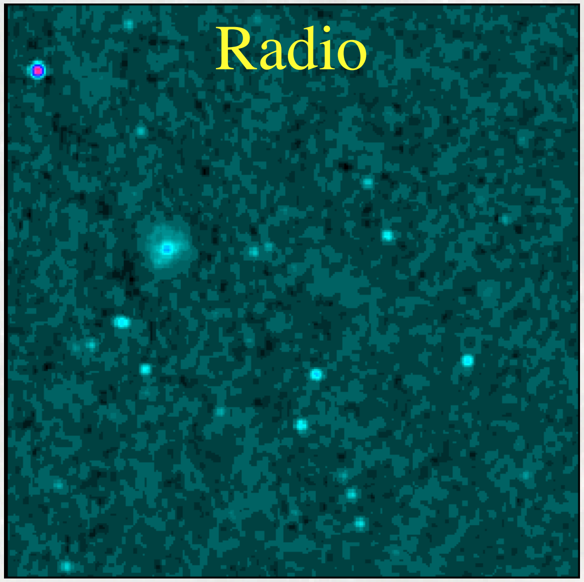
Billions of detected sources, hundreds of measured attributes per source …

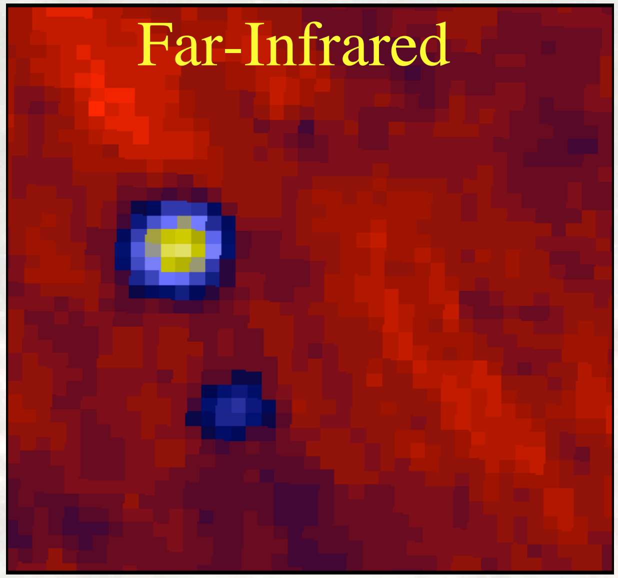


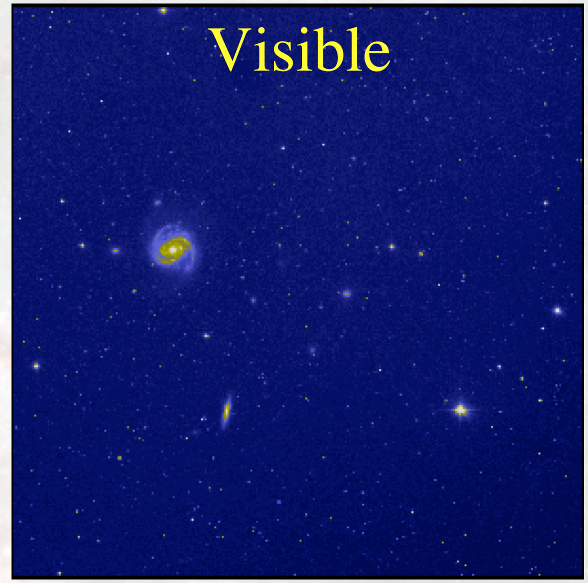
Galactic Center Region (a tiny portion)   2MASS NIR Image

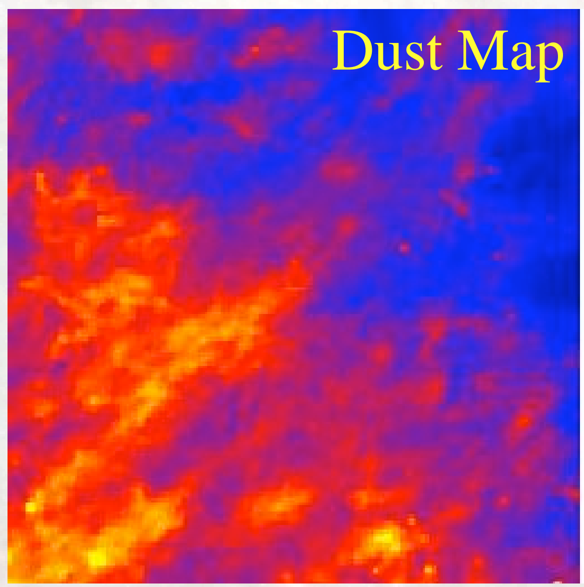
Panchromatic Views of the Universe:

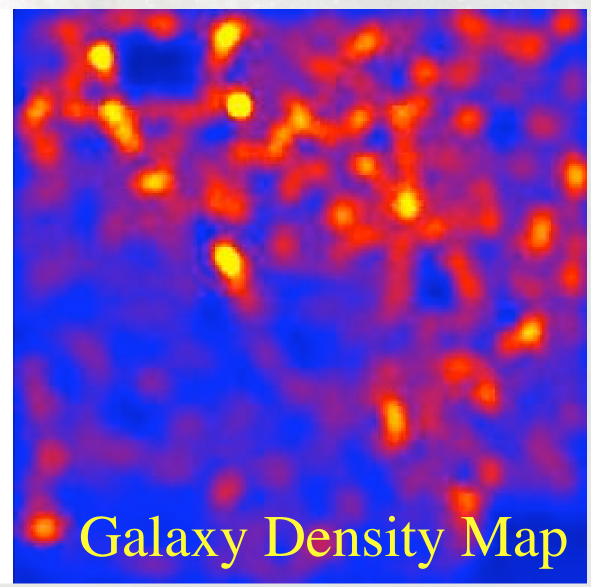
A More Complete, Less Biased Picture











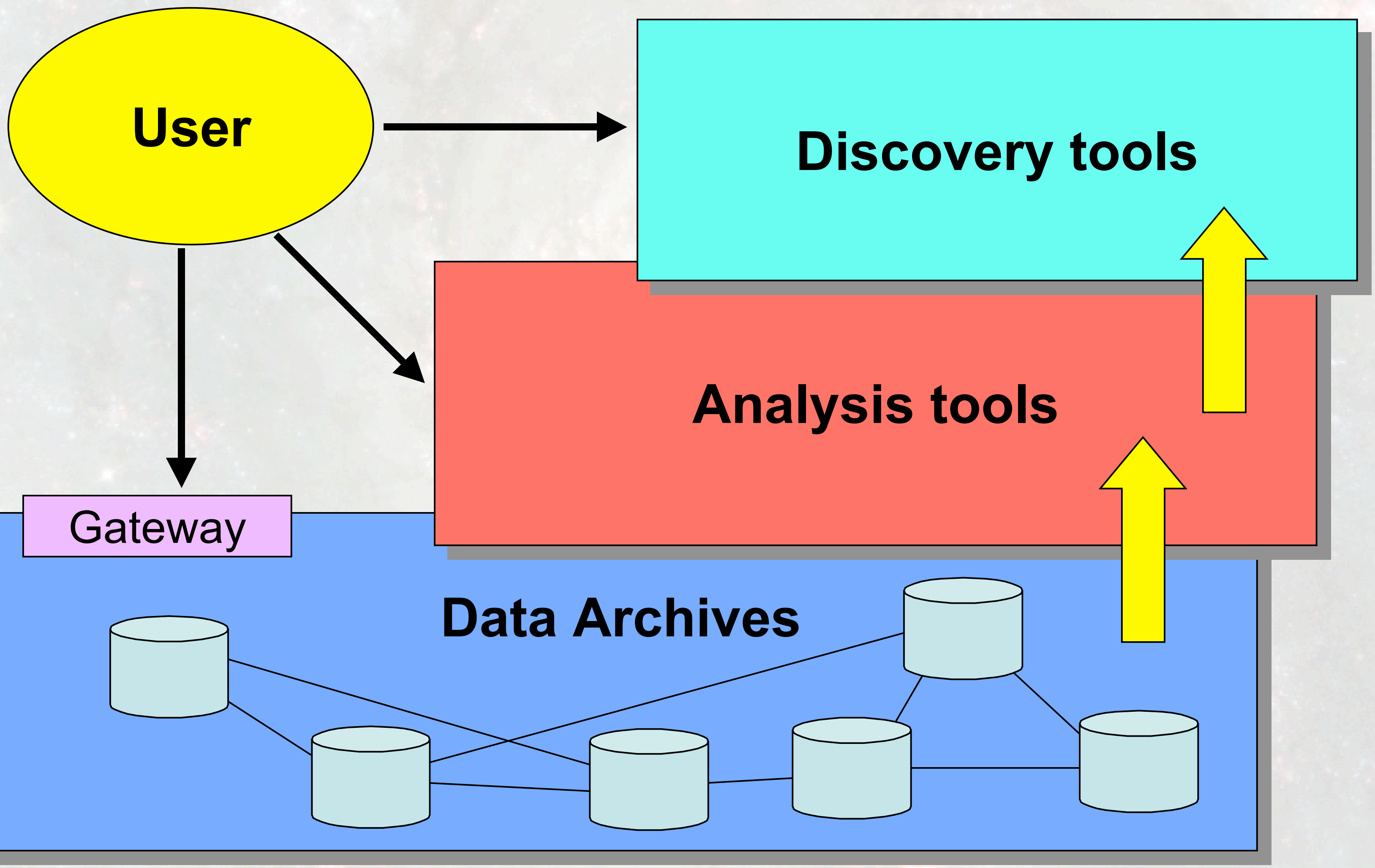
The Changing Face of Observational Astronomy

- **Large digital sky surveys** are becoming the dominant source of data in astronomy: > 100 TB, growing rapidly
  - Spanning many wavelengths, ground- and space-based
  - Also: Digital libraries, Observatory archives
  - Also: Massive numerical simulations
  - Soon: synoptic (multi-epoch or repeated) sky surveys (PB scale)
  - NB: Human Genome is < 1 GB, Library of Congress ~ 20 TB
- **Old style:** studies of individual sources or small samples (~ 10^1 - 10^3 objects), GB-scale data sets
- **New style:** samples of ~ 10^6 - 10^9 sources, TB-scale data sets (soon: PB scale), increasing complexity
- **Data sets many orders of magnitude larger, more complex, and more homogeneous than in the past**

The Virtual Observatory Concept

- Astronomical community response to the scientific and technological challenges posed by massive data sets
  - Highest recommendation of the NAS Decadal Astronomy and Astrophysics Survey Committee ⇨ NVO
  - International growth ⇨ IVOA
- Provide content (data, metadata) services, standards, and analysis/compute services
  - Federate the existing and forthcoming large digital sky surveys and archives, facilitate data inclusion and distribution
  - Develop and provide data exploration and discovery tools
  - ***Technology-enabled, but science-driven***
- A complete, dynamical, distributed, open **research environment for the new astronomy with massive and complex data sets**

VO: Conceptual Architecture



Scientific Roles and Benefits of a VO

- **Facilitate science with massive data sets** (observations and theory/simulations) ⟹ **efficiency amplifier**
- Provide an **added value** from federated data sets (e.g., multi-wavelength, multi-scale, multi-epoch …)
  - Historical examples: the discoveries of Quasars, ULIRGs, GRBs, radio or x-ray astronomy …
- **Enable and stimulate some *new* science** with massive data sets (not just old but bigger)
- **Optimize the use of expensive resources** (e.g., space missions and large ground-based telescopes)
  - Target selection from wide-field surveys
- Provide R&D drivers, application testbeds, and stimulus to the **partnering disciplines** (CS/IT, statistics …)

Broader and Societal Benefits of a VO

- **Professional Empowerment:** Scientists and students anywhere with an internet connection would be able to do a first-rate science $\Rightarrow$ A broadening of the talent pool in astronomy, democratization of the field
- **Interdisciplinary Exchanges:**
  - The challenges facing the VO are common to most sciences and other fields of the modern human endeavor
  - Intellectual cross-fertilization, avoid wasteful duplication
- **Education and Public Outreach:**
  - Unprecedented opportunities in terms of the content, broad geographical and societal range, for all educational levels
  - Astronomy as a magnet for the CS/IT education
  - Creating a new generation of science and technology leaders

"Weapons of Mass Instruction"

http://virtualsky.org (R. Williams et al.)

Increase map size

Decrease map size

Show Marker

Clicking on sky will:

- Zoom In
- Zoom Out
- Recenter
- Galaxy lookup
- Star lookup

You are at:

Right Ascension

14.679072

Declination

60.933364

(clean) Go Here

Image courtesy of Digital Palomar Observatory Sky Survey

Image center is RA,Dec=14.679072 60.933364

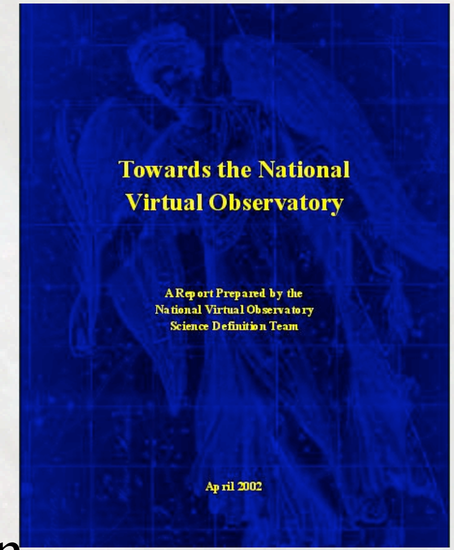
Image width is 53.715627 minutes

Get VS Jpeg Get VS FITS

virtualsky.org

VO Developments and Status

- **In the US:** National Virtual Observatory (NVO)
  - Concept developed by the NVO Science Definition Team (SDT) See the report at ***http://www.nvosdt.org*** ➔
  - NSF/ITR funded project: ***http://us-vo.org***
  - Other, smaller projects under way

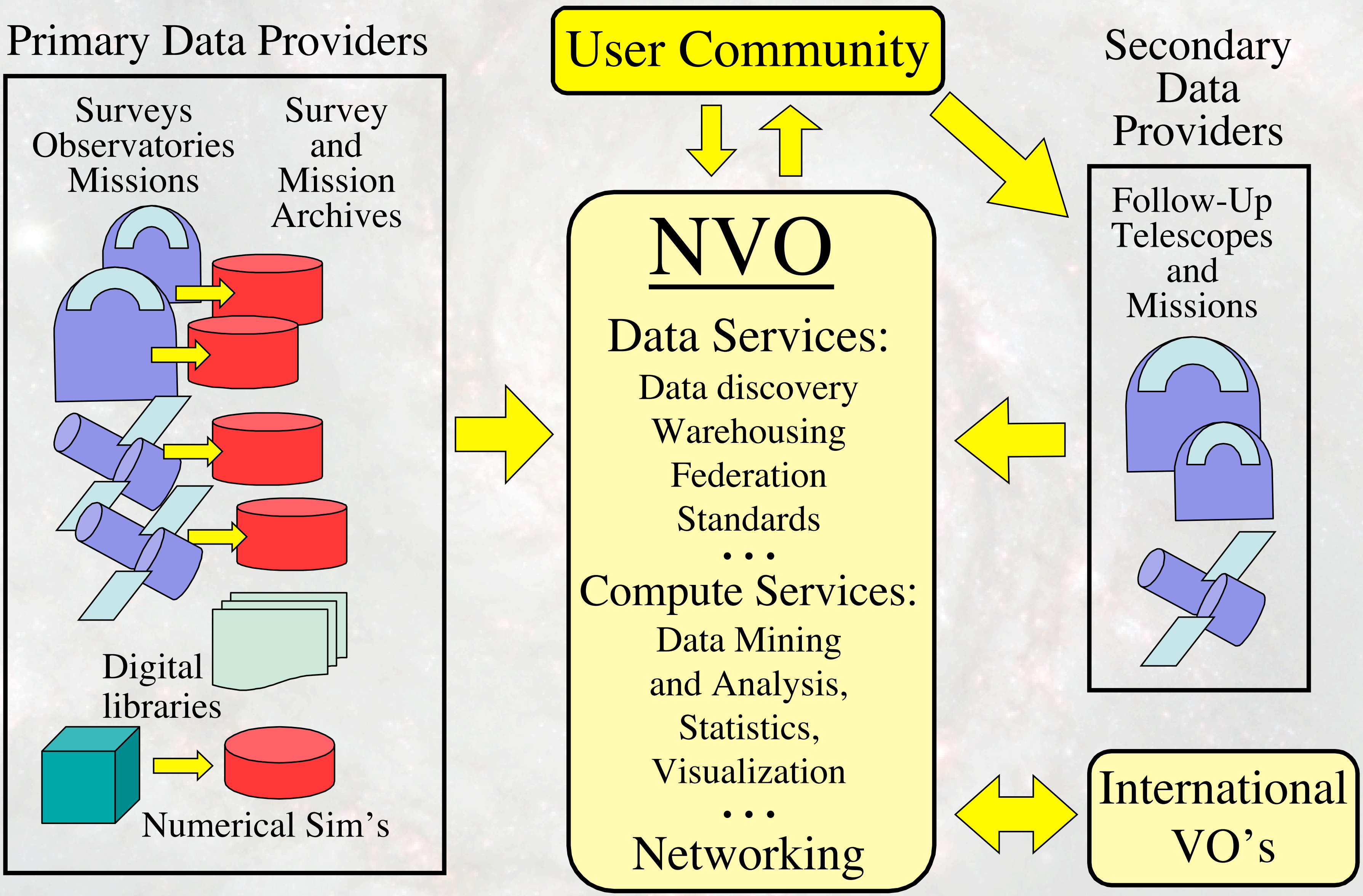


- **Worldwide** efforts:
  - European union: Astrophysical V.O. (AVO)
  - UK: Astrogrid
  - National VO's in Germany, Russia, India, Japan, …
  - International V.O. Alliance (IVOA) formed
- A good synergy of astronomy and CS/IT
- Good progress on data management issues, a little on data mining/analysis, first science demos forthcoming

The NVO Implementation: Organizational Issues

- The NVO has to fulfill its **scientific and educational mandates** (including the necessary IT developments)
- The NVO has to be:
  - **Distributed:** the expertise and the data are broadly spread across the country
  - **Evolutionary:** responding to the changing scientific needs and the changes in the enabling technologies
  - **Responsive** to the needs and constraints of all of its constituents
- The NVO has to communicate/coordinate with:
  - The funding agencies
  - The astronomical community as a whole
  - The existing data centers, archives, etc.
  - The international efforts (IVOA)
  - Other disciplines, especially CS/IT

A Schematic View of the NVO

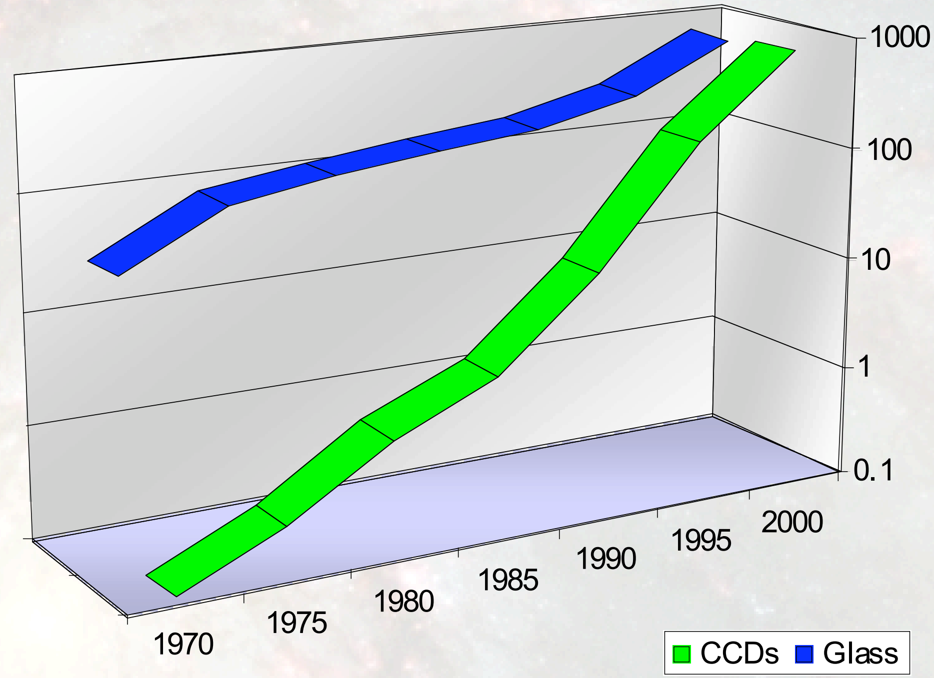


The NVO Organization and Management

- The NVO is ***not*** yet another data center, archive, mission, or a traditional project ⇨ ***It does not fit into any of the usual structures today***
  - It ***transcends the traditional boundaries*** between different wavelength regimes, agency domains (e.g., NSF / NASA)
  - It has an ***unusually broad range of constituents*** and interfaces, and is inherently distributed
  - It requires a ***good inter-agency cooperation***, and a long-term stability of structure and funding
- The NVO represents <u>***a novel type of a scientific organization***</u> for the era of information abundance
- Designing the NVO organizational/management structure is thus *a* ***creative challenge*** *in itself*

Data → Knowledge ?

The exponential growth of data volume (and also complexity, quality) driven by the exponential growth in information technology …



… But our understanding of the universe increases much more slowly -- ***Why?***

- ✦ Methodological bottleneck ➙ VO is the answer
- ✦ Maybe because $S = k \log N$?
- ✦ Human wetware limitations …

➙ AI-assisted discovery ➙ NGVO?

How and Where are Discoveries Made?

- **Conceptual Discoveries:** e.g., Relativity, QM, Strings, Inflation … *Theoretical, may be inspired by observations*
- **Phenomenological Discoveries:** e.g., Dark Matter, QSOs, GRBs, CMBR, Extrasolar Planets, Obscured Universe … *Empirical, inspire theories, can be motivated by them*

New Technical Capabilities $\xrightarrow{\text{IT/VO}}$ Observational Discoveries $\xleftrightarrow{\text{(VO)}}$ Theory

Phenomenological Discoveries:

- Pushing along some parameter space axis ⟸ VO useful
- Making new connections (e.g., multi-□) ⟸ **VO critical!**

Understanding of complex (astrophysical) phenomena requires complex, information-rich data (and simulations?)

The VO-Enabled, Information-Rich Astronomy for the 21st Century

- Technological revolutions as the drivers/enablers of the bursts of scientific growth
- Historical examples in astronomy:
  - 1960's: the advent of electronics and access to space

    Quasars, CMBR, x-ray astronomy, pulsars, GRBs, …
  - 1980's - 1990's: computers, digital detectors (CCDs etc.)

    Galaxy formation and evolution, extrasolar planets, CMBR fluctuations, dark matter and energy, GRBs, …
  - 2000's and beyond: information technology

The next golden age of discovery in astronomy?

Some Musings on CyberScience

- Enable a broad spectrum of users/contributors
  - From large teams to small teams to individuals
  - Data volume ~ Team size
  - Scientific returns $\neq f(\text{team size})$
- Transition from data-poor to data-rich science
  - Chaotic ➔ Organized … However, *some* chaos (or the lack of excessive regulation) is good, as it correlates with the creative freedom (recall the WWW)
- Computer science as the "new mathematics"
  - It plays the role in relation to other sciences which mathematics did in ~ 17th - 20th century

    (The frontiers of mathematics are now elsewhere…)

Concluding Comments and Questions

- Converting new, massive, complex data sets into the knowledge and understanding is ***a universal problem facing all sciences today***
- Quantitative changes in data volumes + IT advances:
  ➞ ***Qualitative changes in the way we do science***
- (N)VO is an example of ***a new type of a scientific research environment*** dealing with such challenges and opportunities
- This requires ***new types of scientific management and organization structures***, a challenge in itself
- The real intellectual challenges are methodological: how do we formulate ***genuinely new types of scientific inquiries, enabled by this technological revolution?***