Big Mirrors, Bayesian Evangelists and the Public:

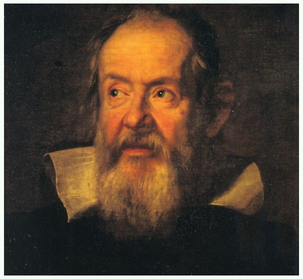
How Advances in Mirror Technology, detectors, databases, machine learning and crowd-sourcing is driving Astronomy into the future

Michael Way (NASA/Goddard Institute for Space Studies)

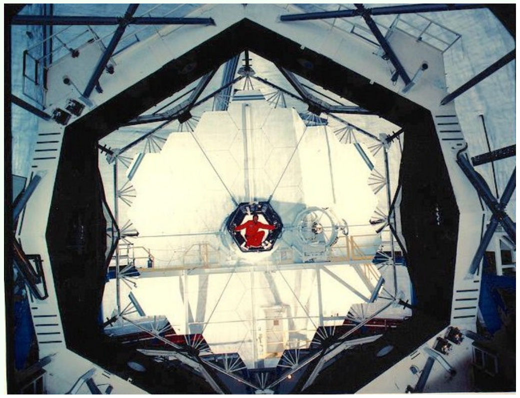
Outline

- **Mirrors**: Single dish to segments
- **Detectors**: paving with plates or silicon?
- **Data**: From ascii files to databases
- **Crowd-sourcing**: AAVSO → Zooniverse
- **Machine Learning**
- **Examples**…

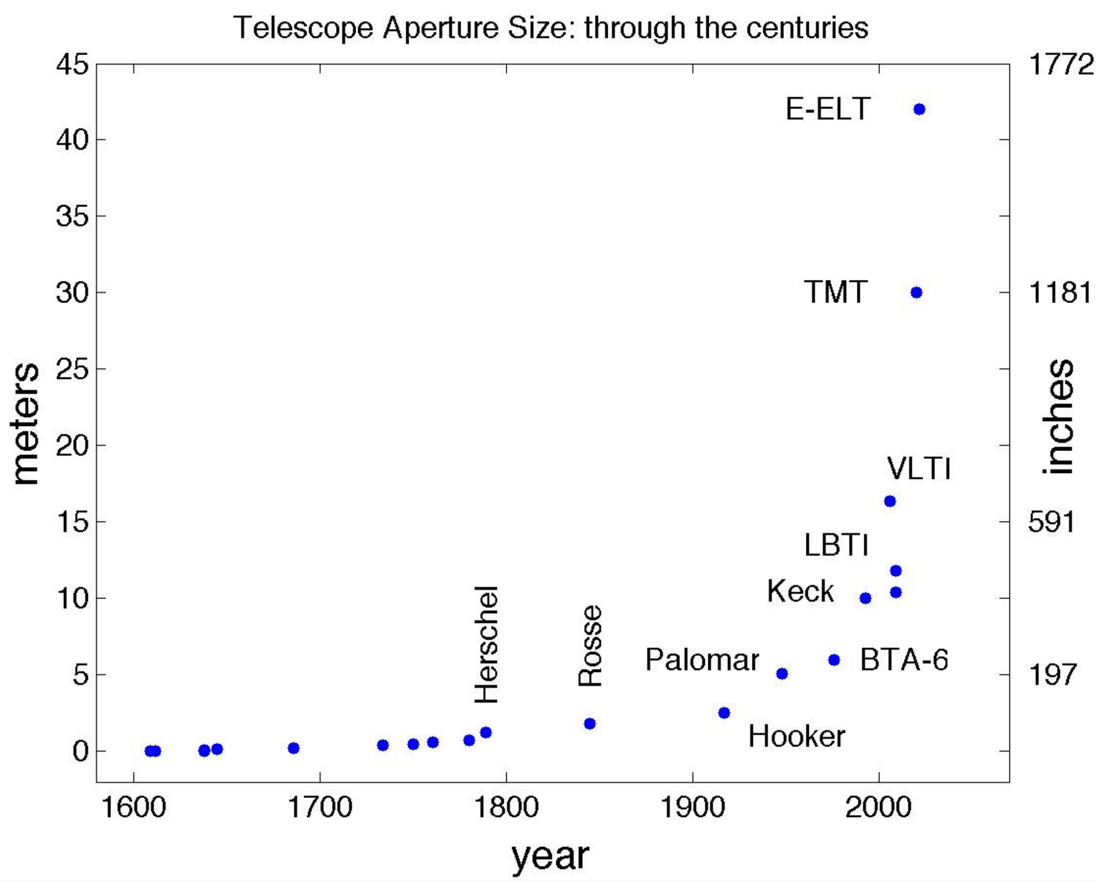
Mirrors: single to segments

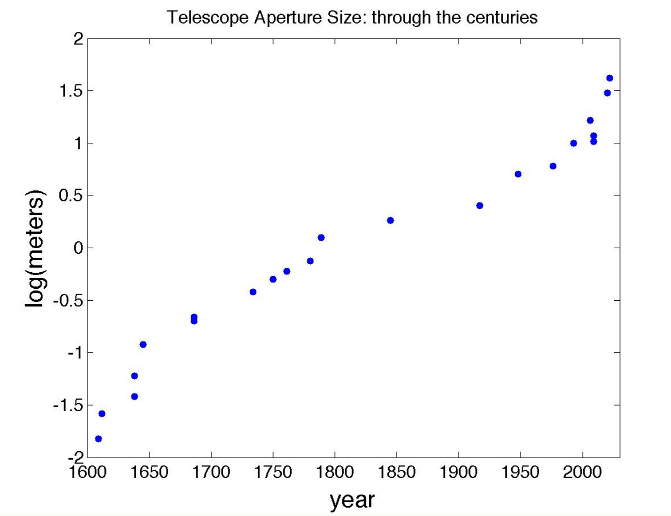


- Lippershey/Galileo refractor: 1.5cm? 1608/1609
- Newton/Hooke reflector: 3.3cm/18cm: 1668/1674
- Herschel: 1.26m: 1789
- Leviathan/Hooker: 1.83m/2.54m: 1845/1917
- Hale/BTA: 5.08m/6m: 1948/1976
- Keck/GTC: 10m/10.4m: 1993/2009
- TMT/E-ELT: 30m/42m: 2020?



Mirror Technology



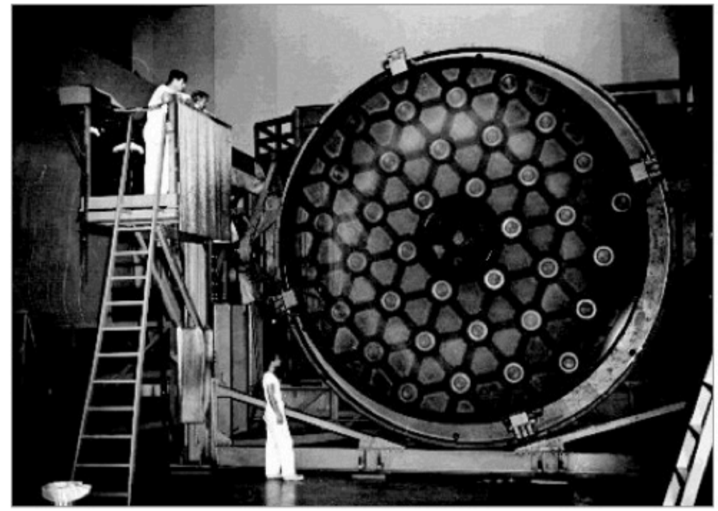


Telescopes/Culture

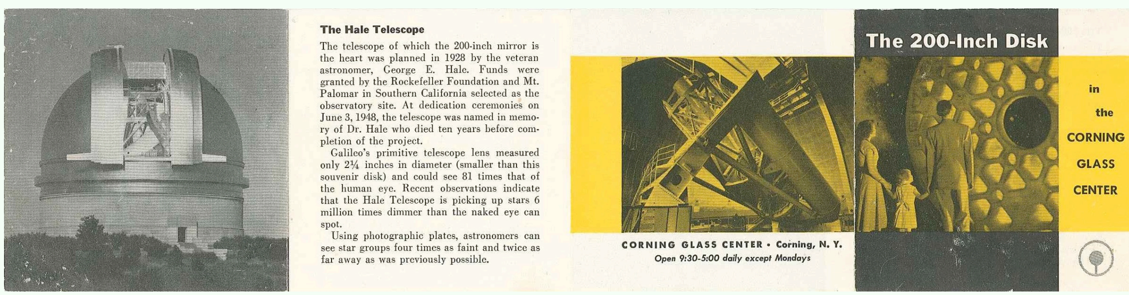
Oct. 3, 1947: Birth of Palomar's 'Giant Eye'

By Tony Long | October 3, 2011 | 6:30 am | Categories: 20th century, Astronomy, Engineering

1947: After 13 years of grinding and polishing, the Palomar Observatory mirror is completed at Caltech.



It was, at the time, the largest telescope mirror ever made in the United States, measuring 200 inches in diameter. Following its completion, the disk was mounted in Palomar's Hale Telescope and first used in January 1949 to take pictures of the Milky Way. Edwin Hubble was the first astronomer to make images using the new scope.



The Hale Telescope

The telescope of which the 200-inch mirror is the heart was planned in 1928 by the veteran astronomer, George E. Hale. Funds were granted by the Rockefeller Foundation and Mt. Palomar in Southern California selected as the observatory site. At dedication ceremonies on June 3, 1948, the telescope was named in memory of Dr. Hale who died ten years before completion of the project.

Galileo's primitive telescope lens measured only 2¼ inches in diameter (smaller than this souvenir disk) and could see 81 times that of the human eye. Recent observations indicate that the Hale Telescope is picking up stars 6 million times dimmer than the naked eye can spot.

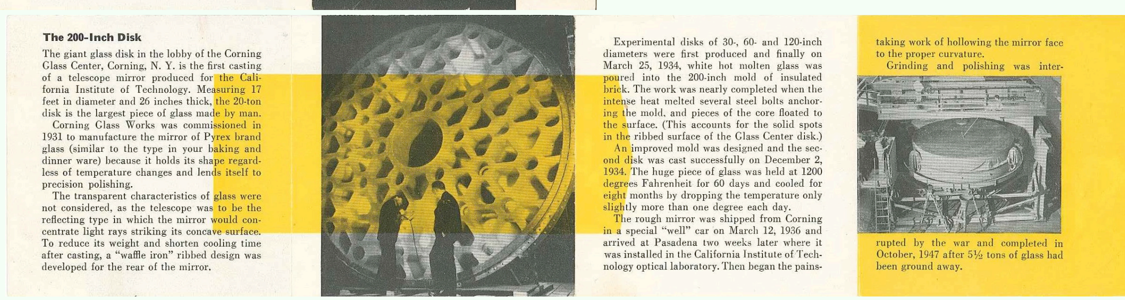
Using photographic plates, astronomers can see star groups four times as faint and twice as far away as was previously possible.

CORNING GLASS CENTER • Corning, N. Y.

Open 9:30-5:00 daily except Mondays

The 200-Inch Disk

in the CORNING GLASS CENTER



The 200-Inch Disk

The giant glass disk in the lobby of the Corning Glass Center, Corning, N. Y. is the first casting of a telescope mirror produced for the California Institute of Technology. Measuring 17 feet in diameter and 26 inches thick, the 20-ton disk is the largest piece of glass made by man.

Corning Glass Works was commissioned in 1931 to manufacture the mirror of Pyrex brand glass (similar to the type in your baking and dinner ware) because it holds its shape regardless of temperature changes and lends itself to precision polishing.

The transparent characteristics of glass were not considered, as the telescope was to be the reflecting type in which the mirror would concentrate light rays striking its concave surface. To reduce its weight and shorten cooling time after casting, a "waffle iron" ribbed design was developed for the rear of the mirror.

Experimental disks of 30-, 60- and 120-inch diameters were first produced and finally on March 25, 1934, white hot molten glass was poured into the 200-inch mold of insulated brick. The work was nearly completed when the intense heat melted several steel bolts anchoring the mold, and pieces of the core floated to the surface. (This accounts for the solid spots in the ribbed surface of the Glass Center disk.)

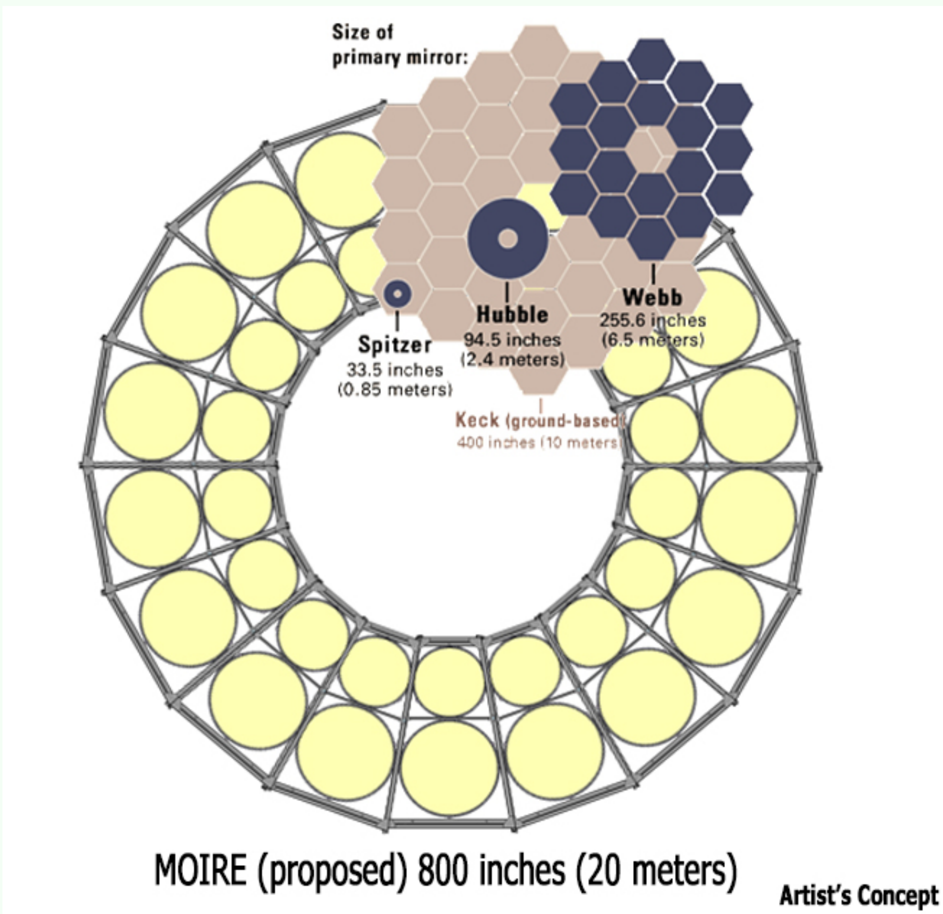
An improved mold was designed and the second disk was cast successfully on December 2, 1934. The huge piece of glass was held at 1200 degrees Fahrenheit for 60 days and cooled for eight months by dropping the temperature only slightly more than one degree each day.

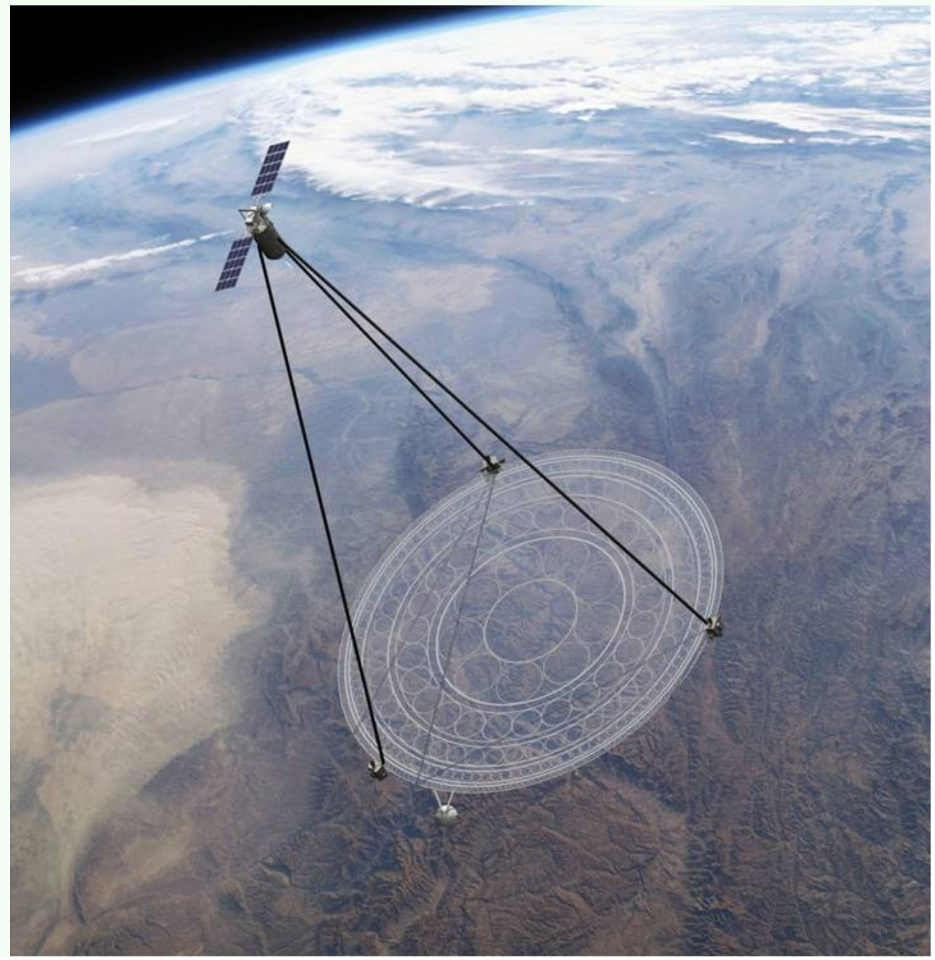
The rough mirror was shipped from Corning in a special "well" car on March 12, 1936 and arrived at Pasadena two weeks later where it was installed in the California Institute of Technology optical laboratory. Then began the painstaking work of hollowing the mirror face to the proper curvature.

Grinding and polishing was interrupted by the war and completed in October, 1947 after 5½ tons of glass had been ground away.

Future: Larger in Space too?

MOIRE: **Membrane Optical** Imager for Real-Time Exploitation





Detectors (Optical)

Eyes, Plates, Photometers and CCDs…

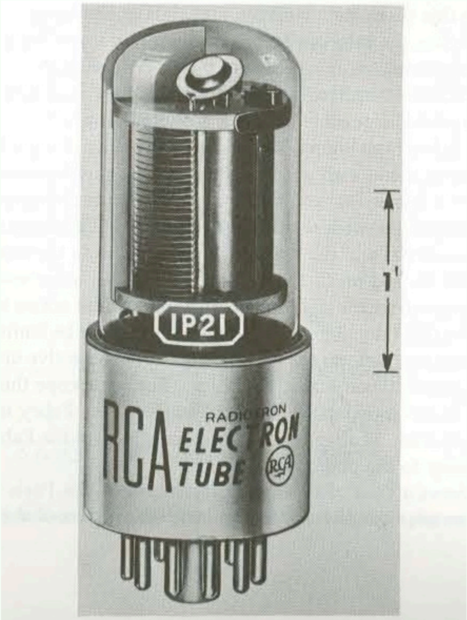


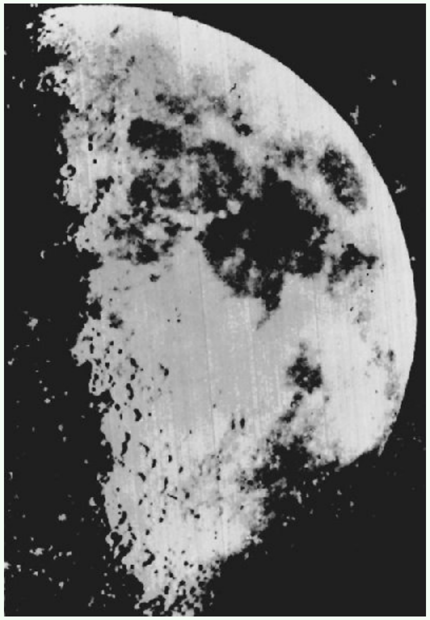
FIG. 13.—RCA 1P21 photomultiplier (see Engstrom 1947)

Eye: QE 1—10%

D=**0.7cm, f=2.2cm, F~3.2**

Upgrades are available

Detectors/QE: Paving the focal plane



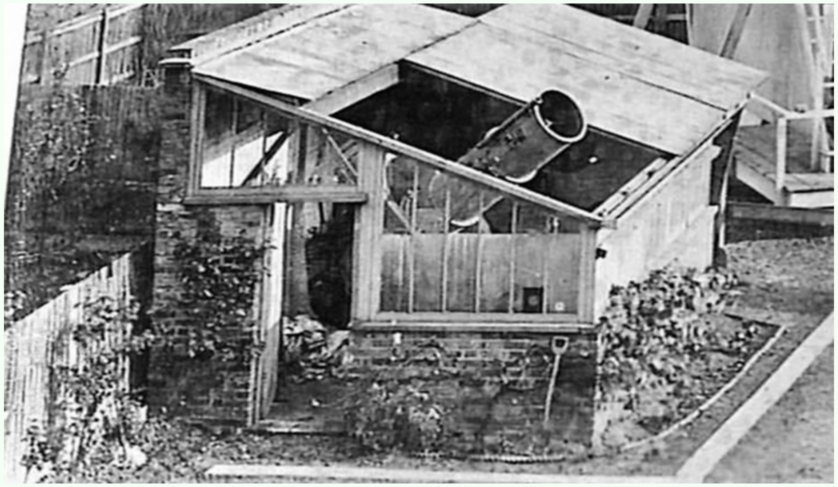
- Niépce (1826) Camera Obscura
- Glass Plates (**Wet**) QE 1%
  - 1839 Daguerre: Moon images
  - 1840 Draper
    - 13cm reflecting telescope
    - Moon (20 min)
  - 1845: Fizeau & Foucault: Sun Images
  - 1850: Bond & Whipple: first star photo - Vega
  - 1872 Miller & Huggins: First spectrograph – Sirius

Earliest extant Daguerre image?: 1851 by John Adam Whipple ---------------------------↑

Detectors/QE: Paving the focal plane

- Glass Plates (**Dry**) QE 1—3%
  - 1876 Huggins (spectrograph)
  - 1880 Henry Draper – first photo of Great Neb. in Orion
  - 1883 Common : First objects fainter than seen by eye?
    - 91cm reflecting/60 min exposure

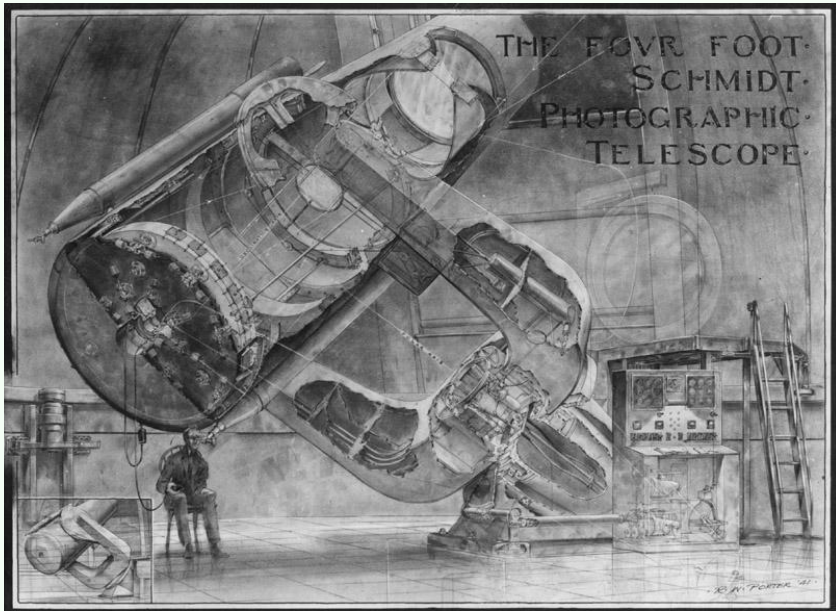
Orion Nebula by Common



Early Common Observatory

Detectors/QE: Paving the focal plane

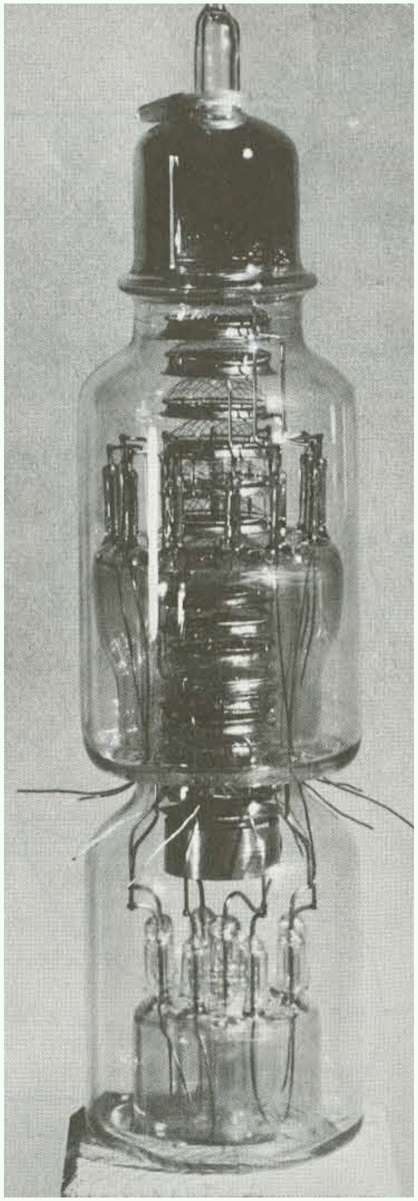
- Large Scale Surveys become feasible with photography
  - 1887 Astrographic Catalogue and Carte du Ciel
    - Aperture ~33 cm, scale: 60 arcsec/mm
    - Field of view: 2°× 2° (Moon diameter ~0.5°)
  - 1948 Oschin Schmidt (Palomar): 1.22m, FoV= 4°×4° degrees
    - 14" x 14" glass plates (41"/mm) – paving the focal plane with glass
  - 1970 DuPont 2.54m (2.1°×2.1°)



Digital(?) Detectors

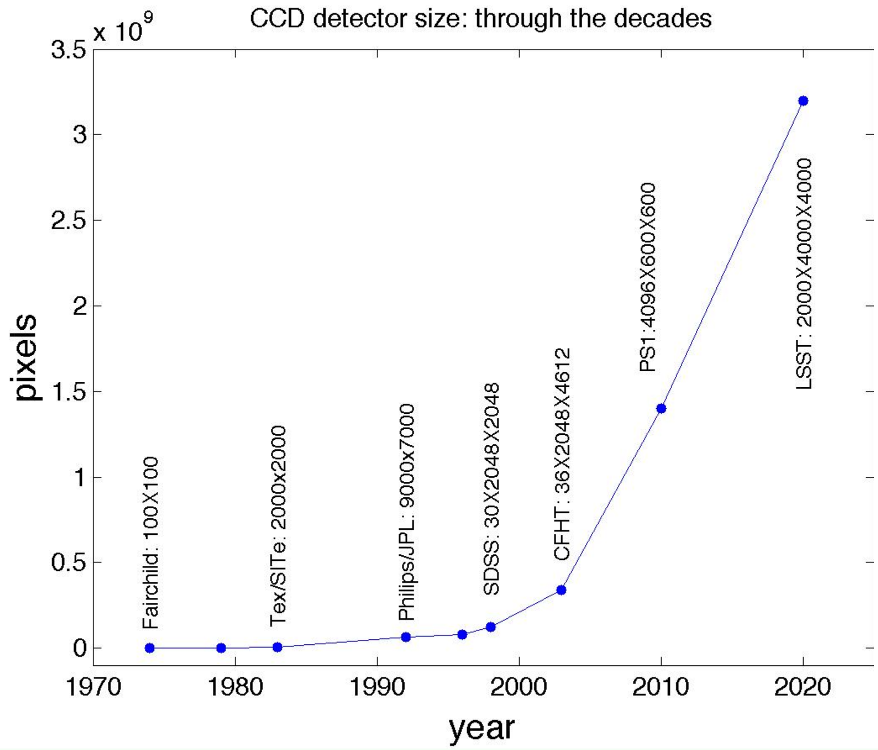
- 1892 – 1980s: Photoelectric photometers
  - Think about it as a single element CCD
  - Digital output? More likely paper tape…
  - X-rays: V2 rockets used in 1949!

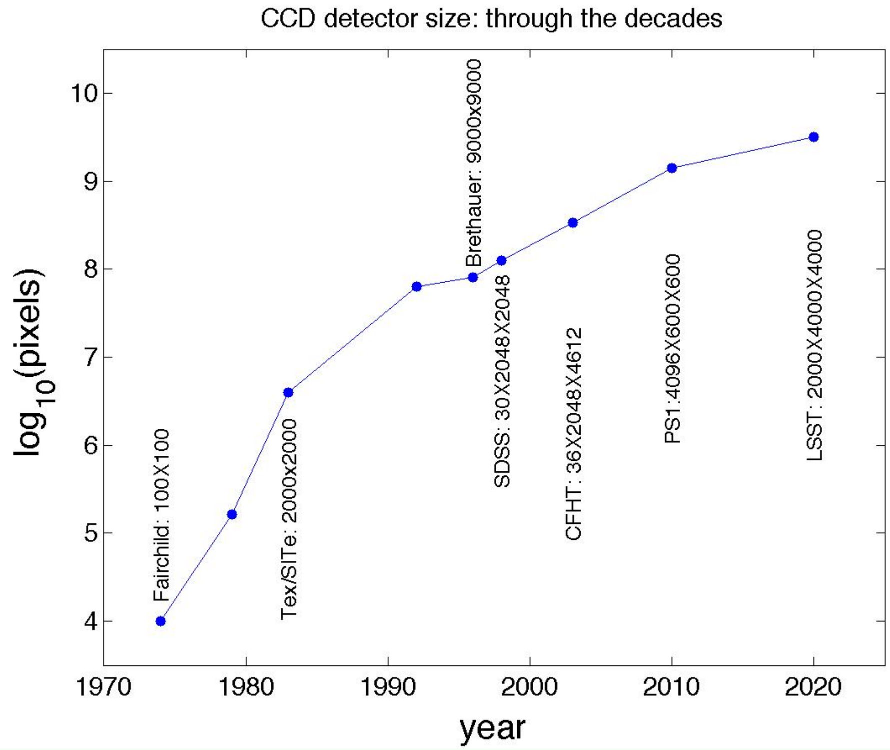
19 stage linear photomultiplier tube developed at the Paris Obs ➔



Detectors/QE: Paving the focal plane

- Charged Coupled Devices (CCDs) QE: 10—90%
  – Boyle and Smith 2009 Nobel Prize for 1970 CCD development at Bell Labs (the 7th from this lab)





Until CCDs were big enough we continued to fill the focal plane with **photographic glass plates** AND **Fibre Spectrographs** (1980):

THE ASTROPHYSICAL JOURNAL, 242:L69–L72, 1980 December 1

MULTIPLE OBJECT SPECTROSCOPY: THE MEDUSA SPECTROGRAPH

JOHN M. HILL, J. R. P. ANGEL, JOHN S. SCOTT, AND DELVIN LINDLEY
Steward Observatory, University of Arizona
AND
PAUL HINTZEN
NASA Goddard Space Flight Center
Received 1980 August 4; accepted 1980 August 28

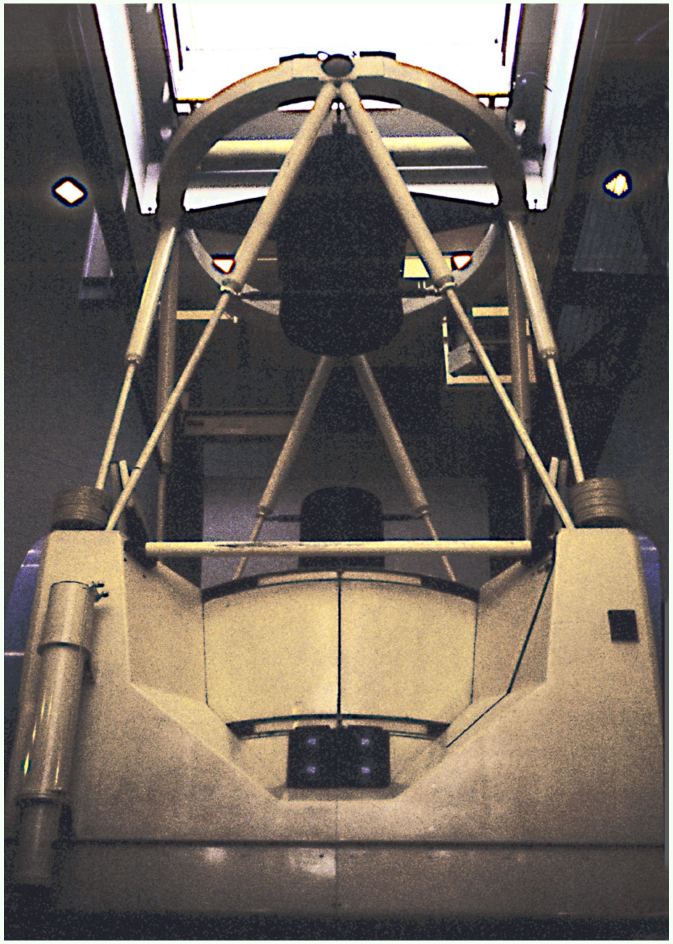
ABSTRACT

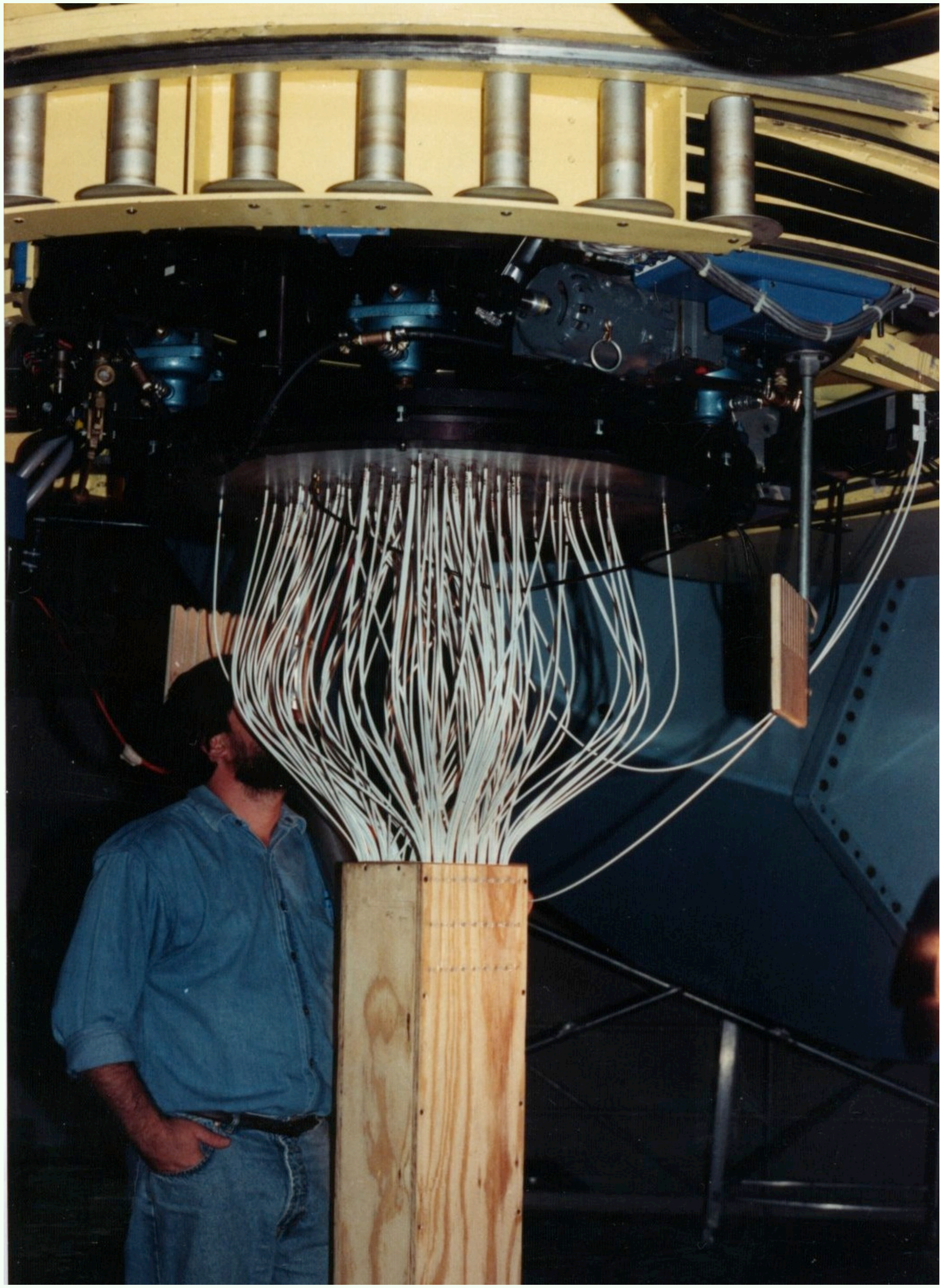
We have built and tested an instrument to obtain simultaneous spectra of many objects in the field of view of the Steward 90 inch (2.29 m) telescope. Short lengths of fused silica fiber 300 μm in diameter are used to bring the light from galaxy images at the Cassegrain focus into a line along the spectrograph slit. From a single exposure of the cluster Abell 1904, which has a redshift of ~20,000 km s^{-1}, we have determined the redshifts of 26 individual galaxies, each with a precision of ~100 km s^{-1}. The present device, while already giving a sixfold reduction in the mean telescope time per galaxy, has significant light losses because it is not ideally matched to the telescope. An instrument being designed for the prime focus will transmit light from each object as efficiently as a conventional spectrograph.

Subject headings: galaxies: redshifts — instruments

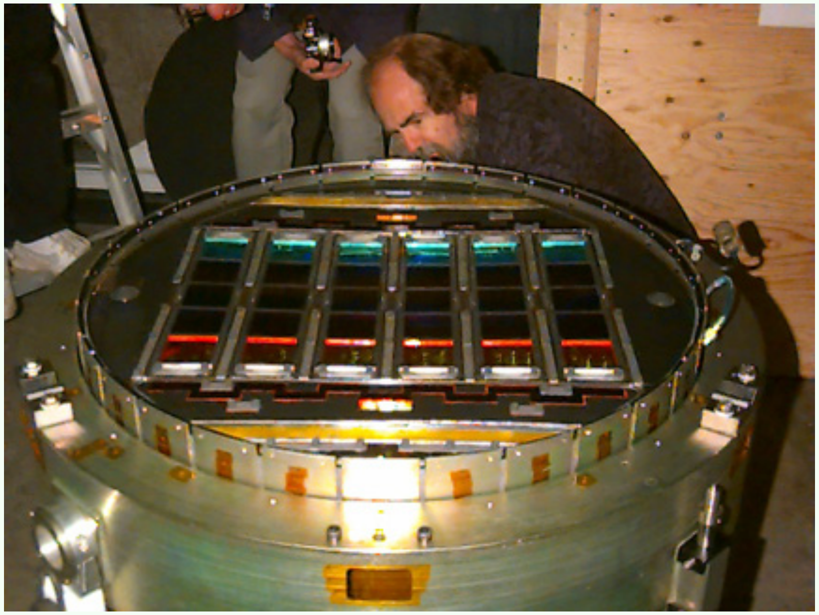
Circa 1990:

Dupont 2.54 m (2° FoV)
Las Campanas, Chile





Sloan Digital Sky Survey Telescope



30 2048x2048 CCDs

600 (1000) fiber spectrograph

Today:

Subaru Prime Focus Spectrograph

2400 fiber positioner, Prime Focus, 1.5 deg FoV

Data:
From ascii files to databases

Revolutions in Data Handling

- Traditionally
  - nearly all data acquisition & data reduction was handled by individual Astronomy researchers
- Today
  - most astronomical data is collected in surveys or by service observations

Part of this has been made possible by the digital and internet revolutions

Media: through the decades…

CCDs (and some photomultipliers) gave us a fully electronic record: Thankfully portable storage was catching up too!

- 9 track tapes [1970] (800-6250 bpi) = 170MB
- TK50/DLT [1984/89 DEC] = 94MB/2.6GB→110GB
- 8mm Exabyte = 112m (2.5/5GB), 54m (1.2/2.4GB)
- 4mm DAT = 90m (2GB), 120m (4GB with DAT2)
- CD = 700MB (100 year lifespan?)
- DVD = 4.7−7GB, BlueRay = 128GB
- Today? It all sits on disk, **or in the cloud**? …

Information storage

Tape rescues big data

Sep 26th 2013, 15:07 by Economist.com

Like 344 Tweet 162

WHEN physicists throw the “on” switch on the Large Hadron Collider (LHC), between three and six gigabytes of data spew out of it every second. That is, admittedly, an extreme example. But the flow of data from smaller sources than CERN, the European particle-research organisation outside Geneva that runs the LHC, is also growing inexorably. At the moment it is doubling every two years. These data need to be stored. And that need for mass storage is reviving a technology which, only a few years ago, seemed destined for the scrapheap: magnetic tape.

LTO-6: 2.5TB, 160MB/s

Distributing Digital Surveys

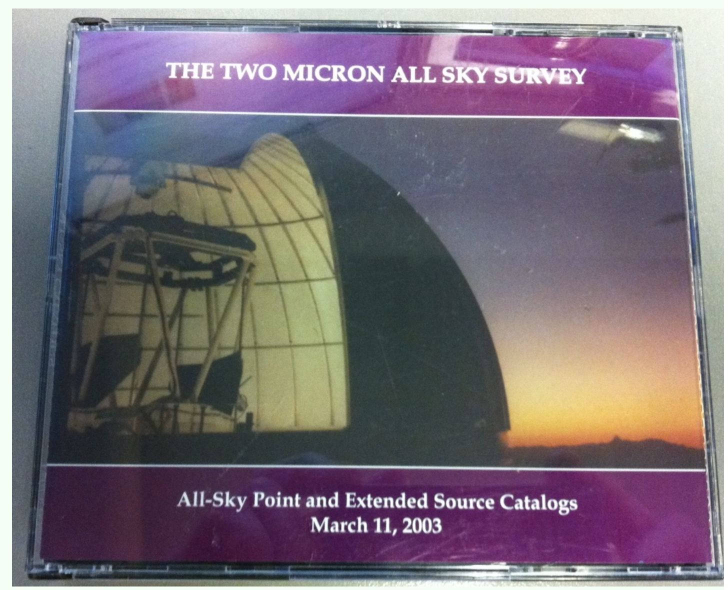
– Digitized Sky Survey 1980s: **102 CDROMs**
– Two Micron All Sky Survey (early 1990s):
  - first large scale fully digital survey (two 1.3m) and catalog.
  - 471 million objects detected
  - Released on **5 double-sided DVDs (43GB)**
  - Full fidelity images ~10TB

1970-80s

Late 1980s/early 1990s

1990s

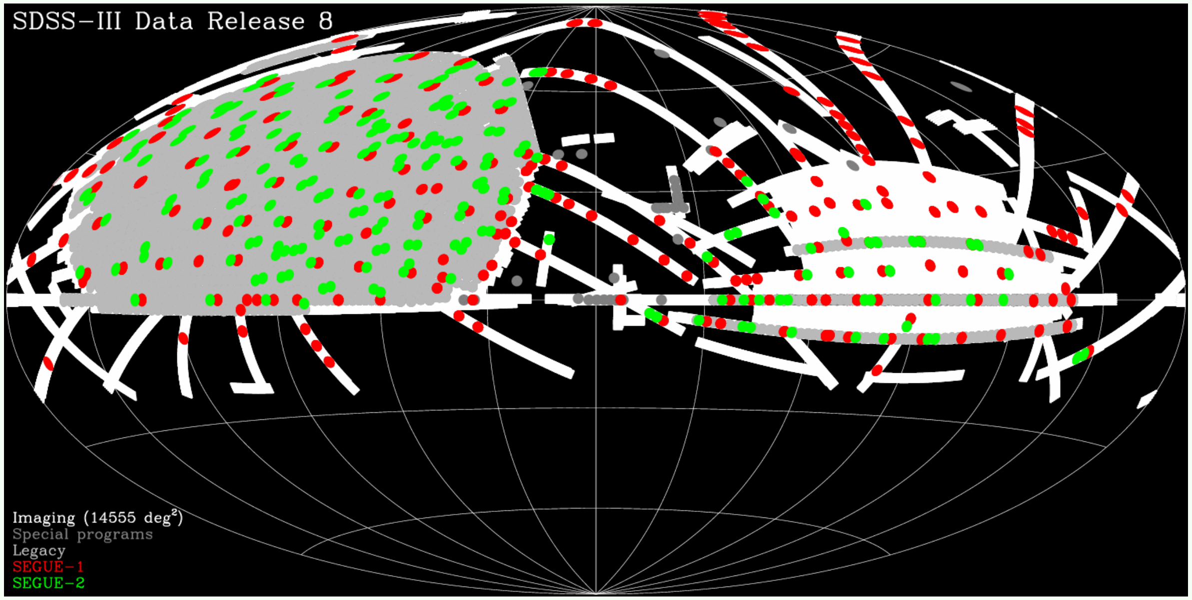
2000s



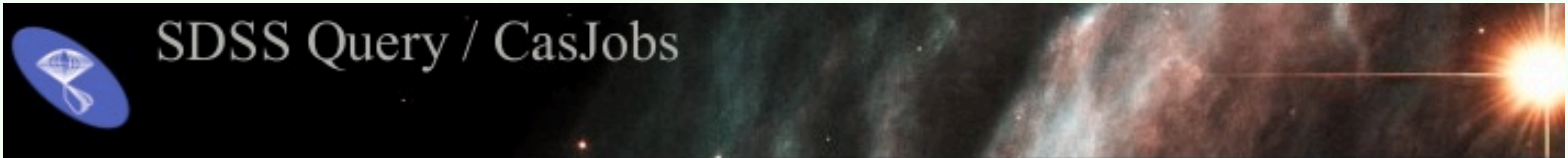
“The Next Generation”

Sloan Digital Sky Survey (2000—Present)

- 120 Megapixel camera, 1.5x1.5 degrees
- 600 (1000) multi-object spectrograph
- ~1 Billion Objects detected, over 1 million spectra



2005 and beyond



Help Tools Query History MyDB Import Groups Output Profile Queues SkyServer Logout

'qso1' Details

Resubmit Job

JobID	TaskName	Context	Queue	Submitted	Started	Finished	Status
5494773	qso1	DR7	600	5/20/2011 3:26:35 PM	5/20/2011 3:26:38 PM	5/20/2011 3:29:40 PM	Finished

Executed on	Rows	Message
DR7Best long	76454	Query Complete

Query

```
Select p.ObjID, p.ra, p.dec,
p.dered_u, p.dered_g, p.dered_r, p.dered_i, p.dered_z, p.petroR50_i, p.petroR90_i,
p.Err_u, p.Err_g, p.Err_r, p.Err_i, p.Err_z, p.petroR50Err_i, p.petroR90Err_i,
s.z, s.zErr, s.zConf, s.zStatus into mydb.qsol from SpecOBJall s, PhotoObjall p
WHERE s.specobjid=p.specobjid
and s.zWarning=0
and (SpecClass=dbo.fSpecClass('QSO') or SpecClass=dbo.fSpecClass('HIZ_QSO'))
and ((flags & 0x8) = 0) and ((flags & 0x2) = 0) and ((flags & 0x40000) = 0)
```

Unexpected Collaborators?

- SDSS dB was built in collaboration with Microsoft
  - Jim Gray + Alex Szalay and others…
  - Interesting large problem, open source data model
  - Still possible to download O(100s GB) data sets
- SciDB: built for next generation data sets (LSST)
  - http://www.scidb.org
  - Not possible to download PB sized data and use it
    - 1PB over 10Gb/s line is 10 days, 1PB = $200 in 2020?
    - I/O not keeping up with other Moore type laws (arXiv:1108.5124v1)
  - R-interface for expert users

Digital Surveys of Today/Tomorrow…

- PanStarrs (2011—2020?)
  - 64x64 array (600x600 CCD) 1.4 Gigapixels
  - ~3TB/night
- Dark Energy Survey (2012—2017)
  - 74 CCDs
  - 1TB/night raw data
- Large Synoptic Survey Telescope (2020—2030)
  - 1PB/night raw data
  - Database size: ~10PB
  - 60PB of images

Crowdsourcing:
From AAVSO to Zooniverse

Crowdsourcing is older than you think?

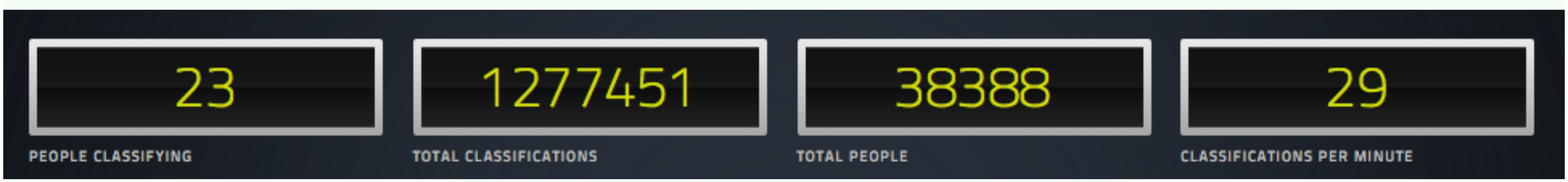
- AAVSO = American Assoc. of Variable Star Observers
  - **Amateur Astronomers contributing observations of variable stars since 1911!**
- 1999: SETI @ Home and copycats
  - Called "Volunteer Computing"
  - Distributed computing, not crowdsourcing

Modern Citizen Science

- 2000: Clickworkers (NASA/Ames)
  - Identifying & Classifying ages of Martian craters from Viking Orbiter images
    - Kanefsky, Barlow and Gulick.
- 2006: Stardust@Home
  - Search aerogel images for tiny dust impacts gathered from tail of Comet Wild

- 2007-Present: GalaxyZoo
  - – Classifying Galaxies in the Sloan Digital Sky Survey
  - – Largest by eye "professional catalog" ~1400 objects
  - – GZ2: 16 million classifications of 304,122 objects
  - – Use Machine Learning to train automated classifiers…
- Zooniverse does a lot more:
  - – Planet Hunting
  - – Transcribe old Weather Logs
  - – Ancient lives (reading old papyri)

- SETILive (Feb 29, 2012) March 10th Economist ⇧



Bayesian Evangelism
and
Machine Learning

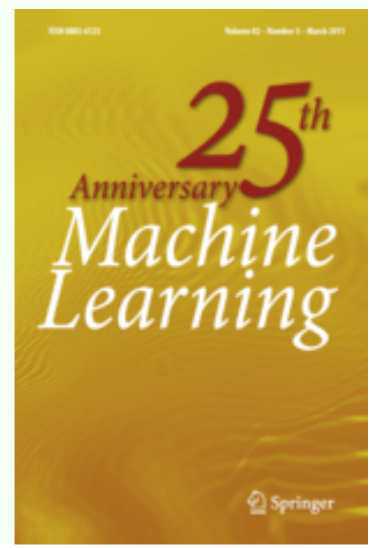
Machine Learning

Rev. Bayes 1702-61

Least Squares

1795 (1809): Gauss
1805: Legendre
1808: Adrian

Shannon (1916-2001)



Machine Learning

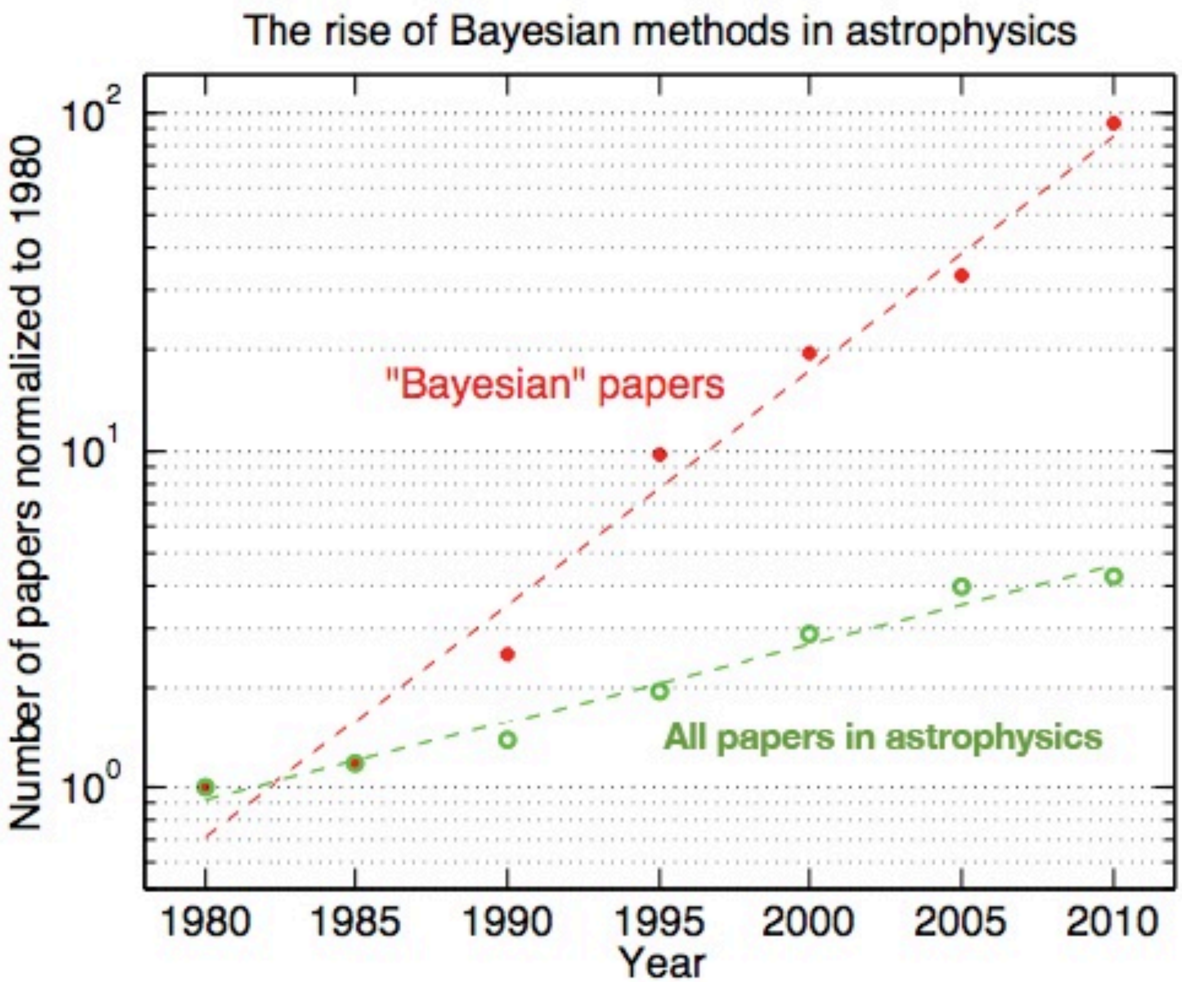
Machine Learning

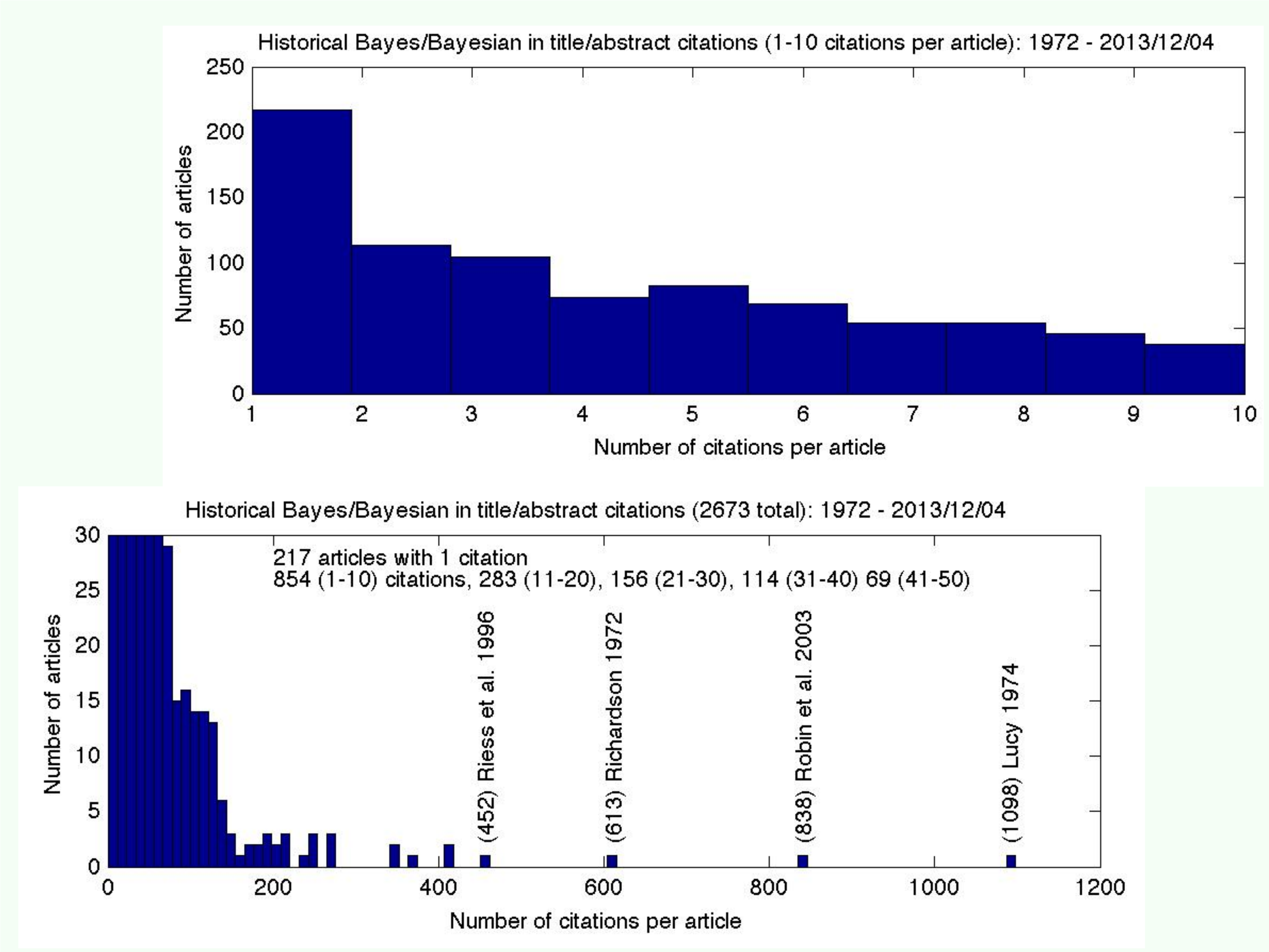
Given new large complex multivariate data
Machine Learning & Data Mining are becoming
more commonly used

- 2600+ plus papers with Bayes in title or abstract
- Large numbers of citations for most popular papers
- Increase in number year by year…

Bayes Evangelism

Review of Bayesian methods in cosmology: Trotta (2008), arxiv: 0803.4089




Historical Bayes/Bayesian in title/abstract citations (1-10 citations per article): 1972 - 2013/12/04
Number of articles
Number of citations per article
Historical Bayes/Bayesian in title/abstract citations (2673 total): 1972 - 2013/12/04
217 articles with 1 citation
854 (1-10) citations, 283 (11-20), 156 (21-30), 114 (31-40) 69 (41-50)
(452) Riess et al. 1996
(613) Richardson 1972
(838) Robin et al. 2003
(1098) Lucy 1974
Number of articles
Number of citations per article

Lucy and Hubble

An iterative technique for the rectification of observed distributions

L. B. Lucy*
Departments of Physics and Astronomy, The University of Pittsburgh, Pittsburgh, Pennsylvania 15213
(Received 15 January 1974; revised 26 March 1974)

A&A 409, 523–540 (2003)
DOI: 10.1051/0004-6361:20031117

Astronomy & Astrophysics

A synthetic view on structure and evolution of the Milky Way

A. C. Robin[1], C. Reylé[1], S. Derrière[2], and S. Picaud[1]

Bayesian-Based Iterative Method of Image Restoration*

William Hadley Richardson
Visibility Laboratory, University of California, San Diego, San Diego, California 92152
(Received 15 September 1970)

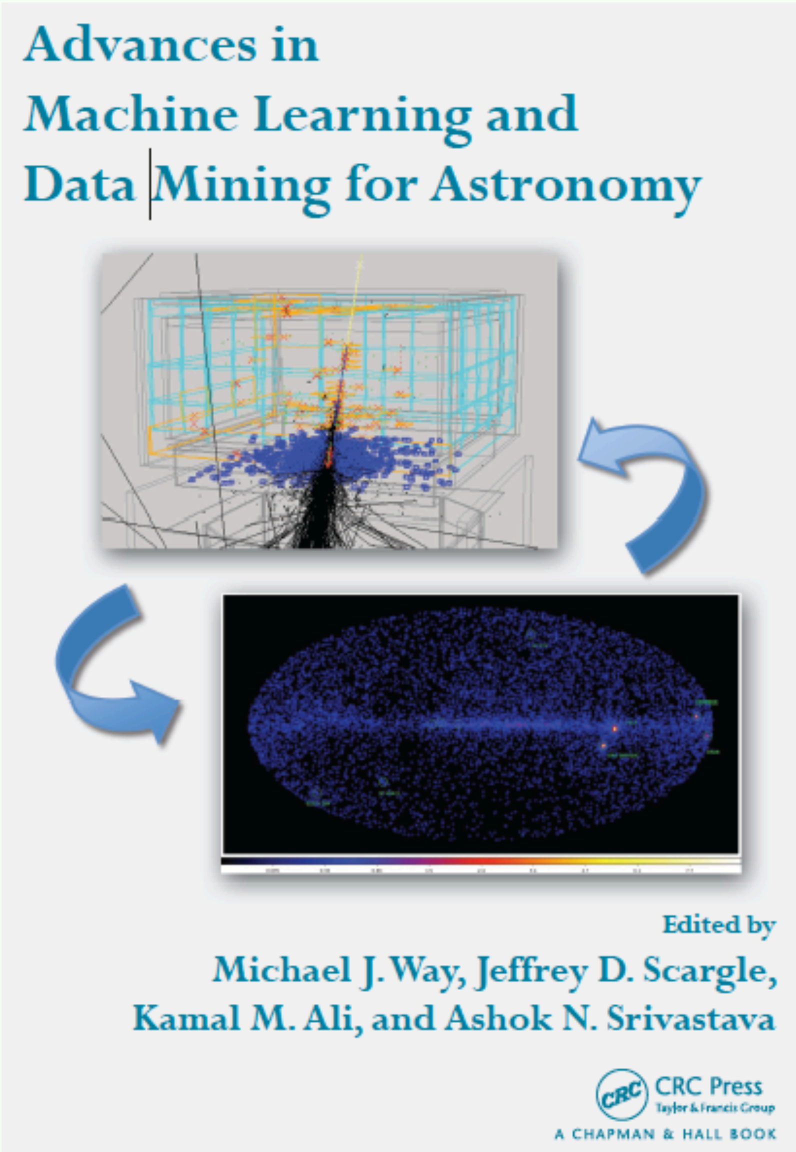
---# of Papers with these in title/abstract (2013/12)---

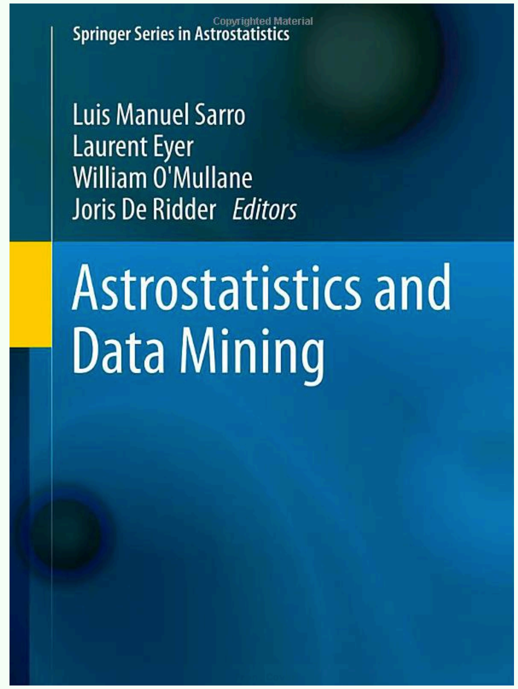
- 1579 "Neural Networks" (well known?) 12288 citations
- 1276 "Data Mining": 7490 citations
- 429 "Machine Learning": 2684 citations
- 125 "Self Organizing Maps" (obscure?) 778 citations

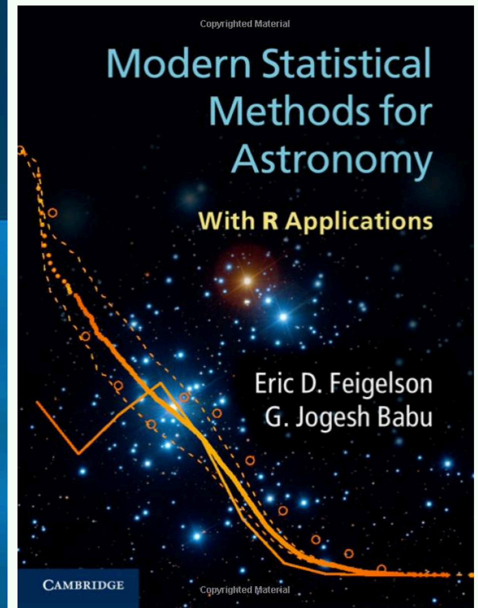
---Some conferences related to Stats/ML in Astronomy---

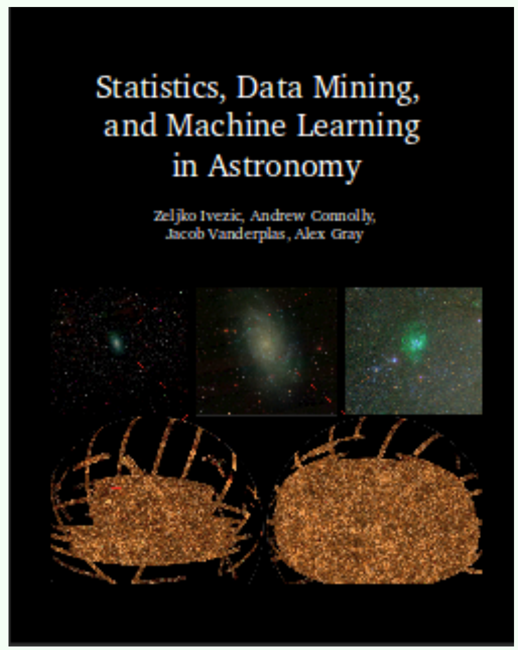
- Statistical Challenges in Modern Astronomy
  - 5 conferences+books 1991—2011, 7 summer schools
- Astrostatistics/Astroinfomatics (https://asaip.psu.edu/)
- SciCoder workshops (http://www.scicoder.org)
- CESS (http://www.giss.nasa.gov/meetings/cess2011)

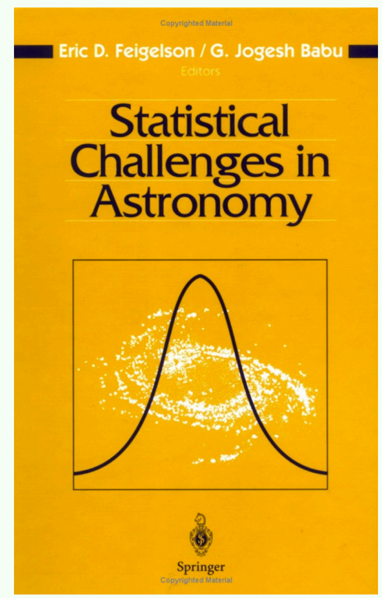
BOOKS











5 volumes!
2003-2013

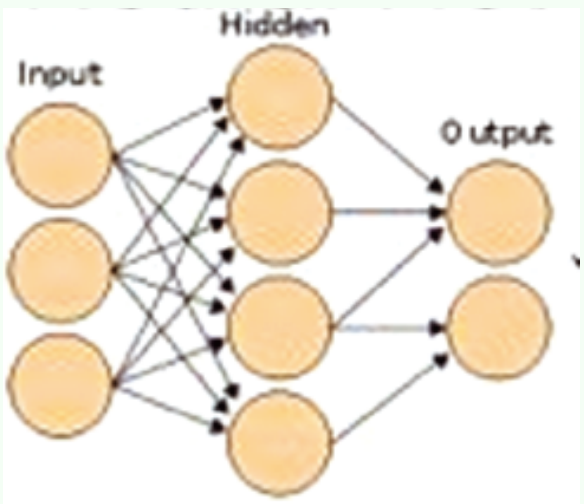
Example from Today?
SDSS+ Crowd-Sourcing + Machine Learning

"Galaxy Zoo: reproducing morphologies via machine learning", Banerji et al. 2010

1. GalaxyZoo morphologies from volunteers (T_{eye})
2. Primary & Secondary Isophotal Parameters (T_{SDSS})

Neural network

Inputs=T_{SDSS}



Outputs=Teye

Primary & Secondary Isophotal Parameters (T_{SDSS})

Table 1. First set of input parameters based on colours and profile fitting.

Name	Description
dered_g-dered_r	$(g - r)$ colour
dered_r-dered_i	$(r - i)$ colour
deVAB_i	de Vaucouleurs fit axial ratio
expAB_i	Exponential fit axial ratio
lnLexp_i	Exponential disc fit log likelihood
lnLdeV_i	de Vaucouleurs fit log likelihood
lnLstar_i	Star log likelihood

Table 2. Second set of input parameters based on adaptive moments.

Name	Description
petroR90_i/petroR50_i	Concentration
mRrCc_i	Adaptive (+) shape measure
aE_i	Adaptive ellipticity
mCr4_i	Adaptive fourth moment
texture_i	Texture parameter

Results

Table 5. Summary of results for the entire sample when using input parameters specified in Tables 1 and 2.

		Galaxy Zoo		
		Early type (per cent)	Spiral (per cent)	Point source/artefact (per cent)
A	Early type	92	0.07	0.6
N	Spiral	0.1	92	0.08
N	Point source/artefact	0.2	0.2	96

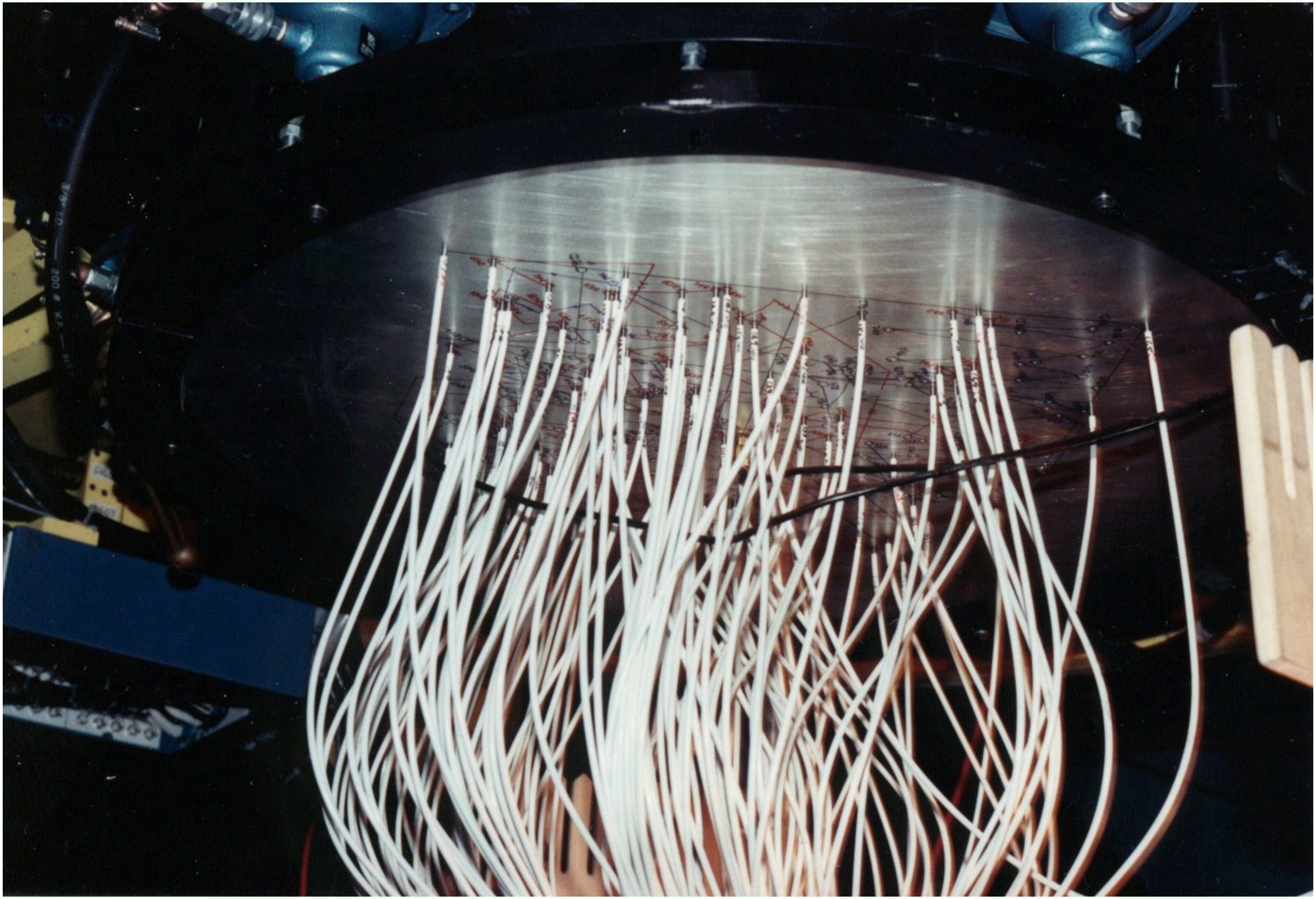
Cultural Changes

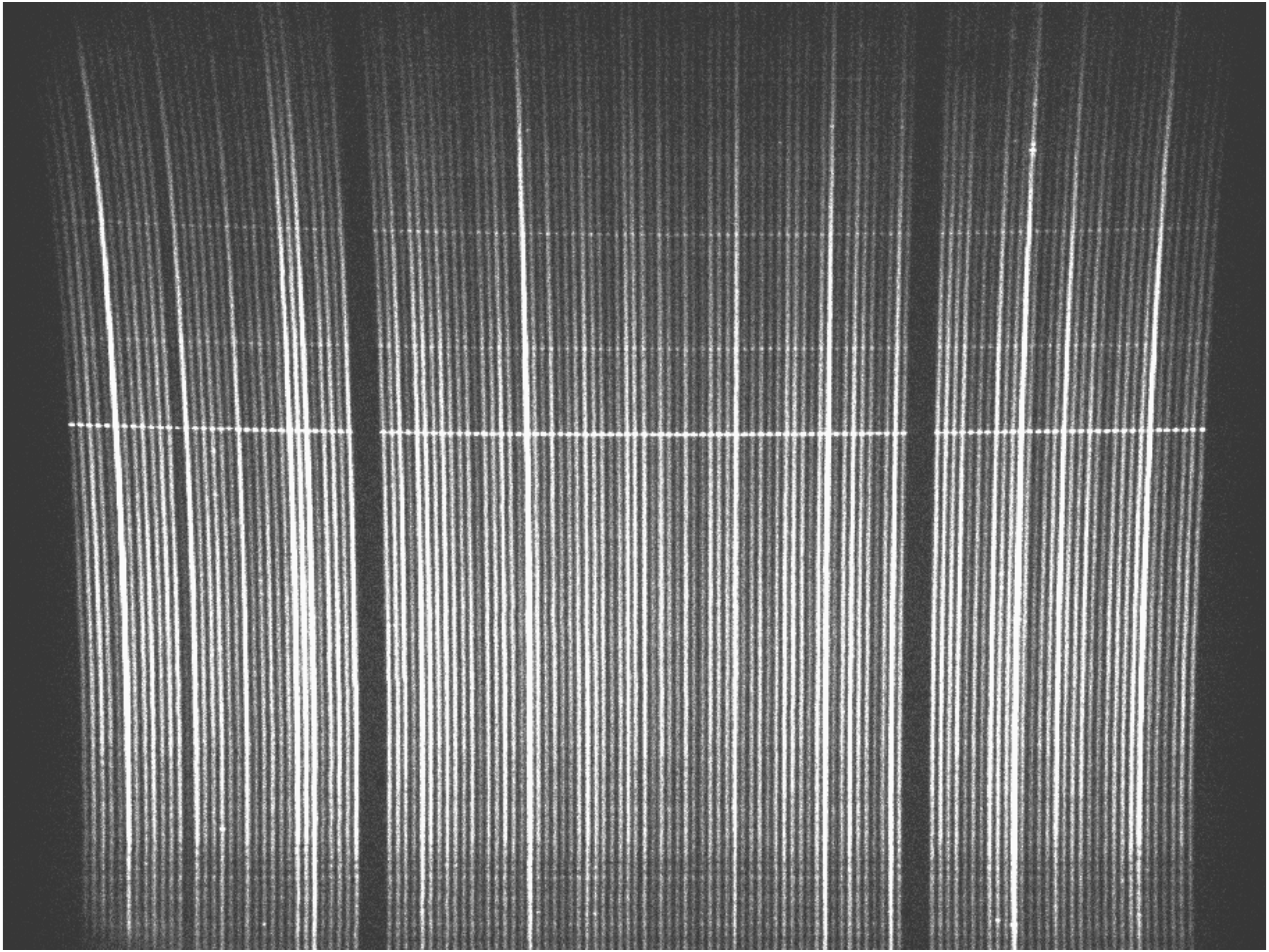
How are these things changing the way we work?

- Avoid applying for grants and telescope time and requisite travel funding
- Avoid specialized instrument knowledge, data acquisition, data reduction, backup/storage concerns
  - Positive **and** Negative aspects
- More time for thinking up good questions?!
  - Submit a query (10 min)
  - Download the data
  - Write your paper

My world… then

- PhD: Spent 2 years in Chile collecting spectroscopic redshifts in (Abell) clusters of galaxies
  - Incredibly specialized instrument knowledge
    - Detector was photon counting, but also imaging like CCD
    - Fiber spectrograph: instrument calibration, sky subtraction, etc.
  - Heavy use of time/labor
    - Plates drilled in Pasadena, shipped to Chile weeks before
    - A 3 night run required showing up 3-4 days before to prepare
      - Marking up plates, checking out instrument, getting to observatory
    - ***very*** long nights: plugging plates beforehand, collecting calibration frames (sky flats) until well after twilight
  - 1—2 years "reducing" the data for thesis





My world… today

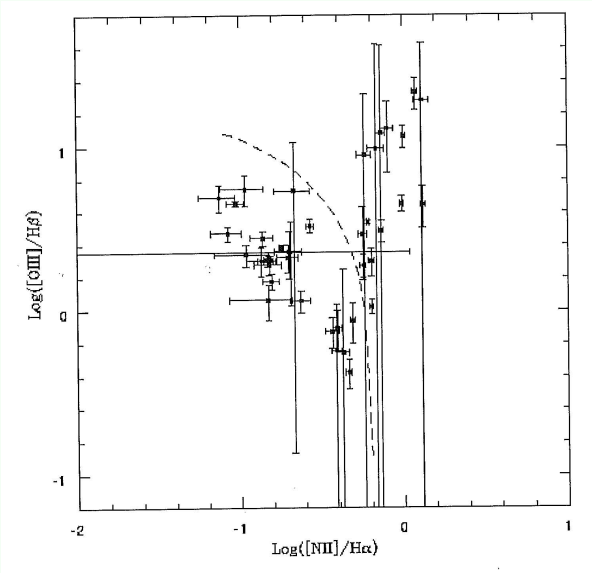
"Galaxy Zoo Morphology & Photometric Redshifts in the Sloan Digital Sky Survey"

1) SDSS Data Release 7 (Oct 2008)
2) GalaxyZoo Data Release 1 (Feb 2011)
   - Morphologies for SDSS galaxies
   - Banerji et al. 2010 isophotal parameters of use
3) Gaussian Process Regression (Foster et al. 2009)
4) Cross-match catalog built in 5 min (March 2011)
5) Paper written in 2 weeks
   - Received March 25, 2011 : Accepted April 21, 2011

Differences in data compression?

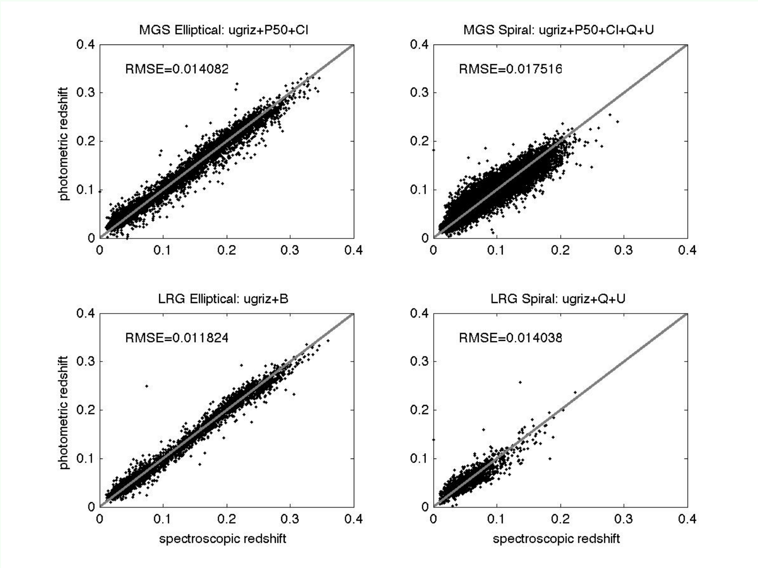
Thesis:

40 DAT (~100GB) + 4 years →



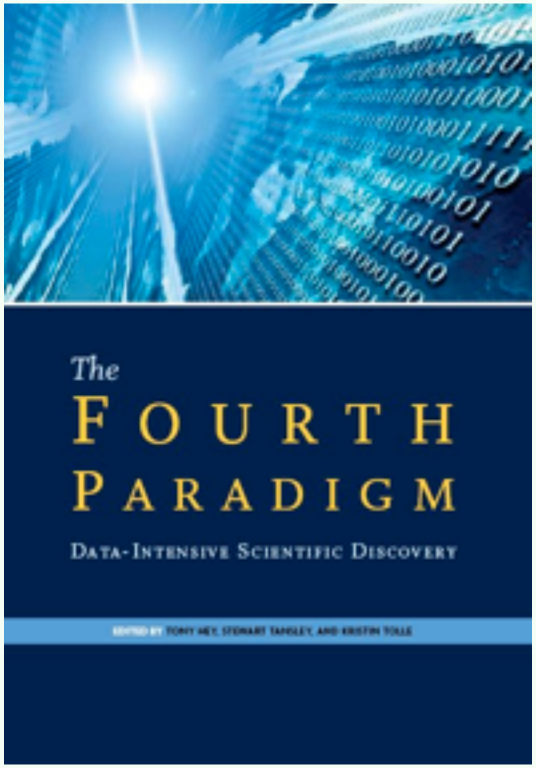
Last Project:

30TB + 1000 years? →



I'm not an Astronomy Chauvinist!

These changes/challenges are not limited to Astronomy



http://www.sciencemag.org/site/special/data

October 2013 in Wired & Quanta

*** 2013/10/02: "A Digital Copy of the Universe"**
https://www.simonsfoundation.org/quanta/20131002-a-digital-copy-of-the-universe-encrypted/

*** 2013/10/03: "Big Data Is Too Big for Scientists to Handle Alone"**
http://www.wired.com/wiredscience/2013/10/big-data-science

*** 2013/10/09: "Scientific Data Has Become So Complex, We Have to Invent New Math to Deal With It"**
http://www.wired.com/wiredscience/2013/10/topology-data-sets

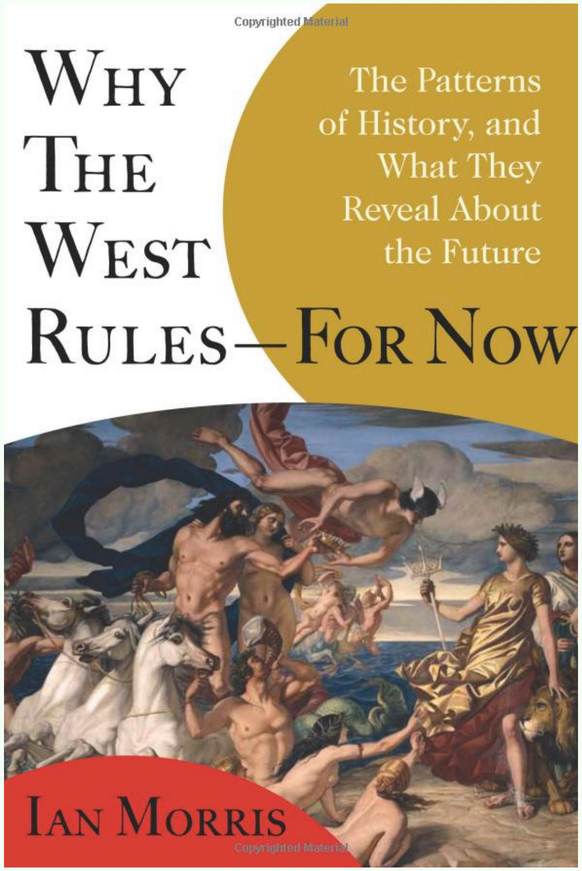
*** 2013/10/11: "Biology's Big Problem: There's Too Much Data to Handle"**
http://www.wired.com/wiredscience/2013/10/big-data-biology

*** 2013/10/14: "How Quantum Computers and Machine Learning Will Revolutionize Big Data"**
http://www.wired.com/wiredscience/2013/10/computers-big-data

Not limited to Natural Science either:

The Humanities

We know **data** is also allowing new collaborations within the humanities & even between science and the humanities:



- History
- Sociology
- Anthropology, Archaeology -- Carbon Dating
- Geology
- Genetics

Lessons for the Business World!?

The Economist

Schumpeter

Titans of innovation

What can business learn from Big Science?

Apr 27th 2013 | From the print edition

AS A technical feat, ATLAS takes some beating. It is the world's biggest microscope, used by physicists at CERN, a large laboratory near Geneva, to probe the fundamental building blocks of matter. Its barrel-shaped body, 45 metres long, 25 metres tall and weighing as much as the Eiffel tower, was assembled in a cavern 100 metres beneath the Swiss countryside from 10m parts, nearly twice as many as in a jumbo jet. It generates more data each day than Twitter does.

Even Foreign Affairs

FOREIGN AFFAIRS

Published by the Council on Foreign Relations

Home | International Editions | Digital Newsstand | Job Board | Account I

Welcome, michael.j.way@gmail.com (Lo

New Issue | Archive | Regions | Topics | Features | Discussions | Video | Books & Reviews | Classroom | About Us

Home › Features › Essays

The Rise of Big Data

How It's Changing the Way We Think About the World

By Kenneth Neil Cukier and Viktor Mayer-Schoenberger

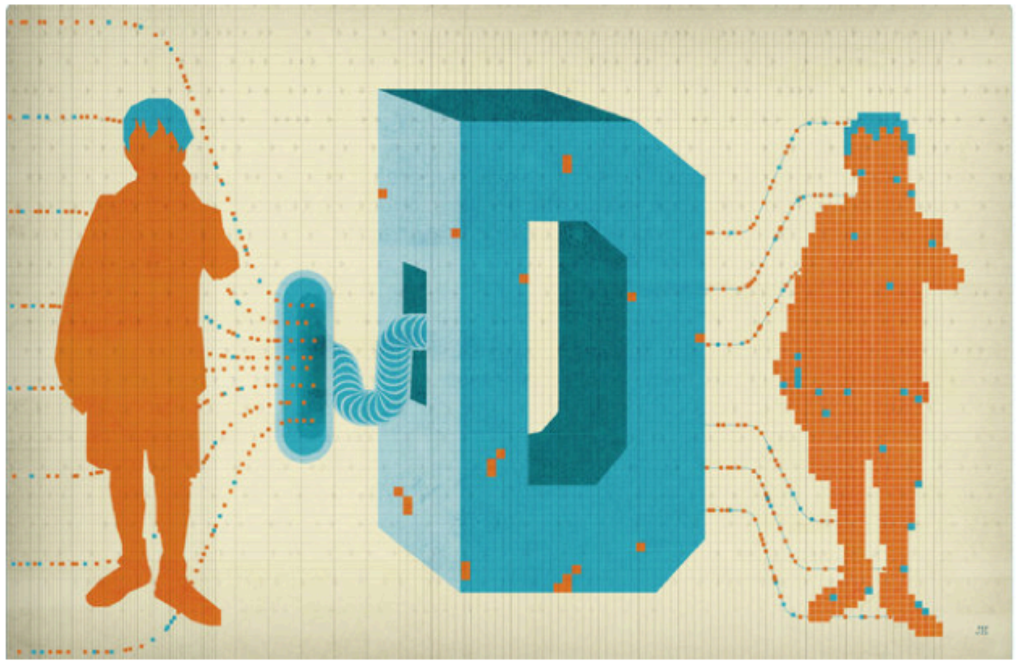
FROM OUR MAY/JUNE 2013 ISSUE

http://www.foreignaffairs.com/articles/139104/
kenneth-neil-cukier-and-viktor-mayer-schoenberger/the-rise-of-big-data

Nor are these changes going to solve all of our problems…

UNBOXED

Sure, Big Data Is Great. But So Is Intuition.



John Hersey

By STEVE LOHR
Published: December 29, 2012

It was the bold title of a conference this month at the Massachusetts Institute of Technology, and of a widely read article in The Harvard Business Review last October: "Big Data: The Management Revolution."

FACEBOOK
TWITTER
GOOGLE+
SAVE

http://www.nytimes.com/2012/12/30/technology/big-data-is-great-but-dont-forget-intuition.html

Conclusions

1. Science is becoming more data intensive
2. We are exploring new collaborations to deal with the data deluge forced upon us by technological advances
3. This also leads to rethinking our methodologies:
   – Probability theory, statistics, machine learning, data mining, etc…
   – And perhaps our productivity?