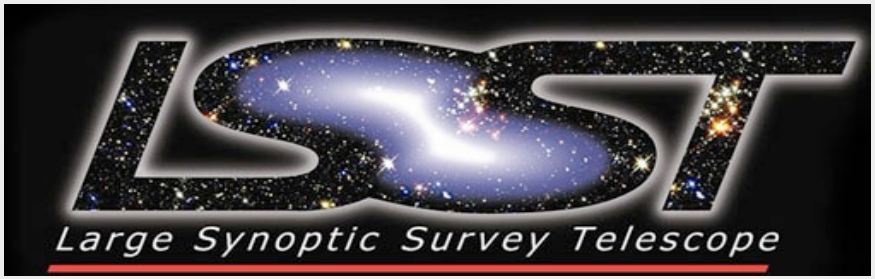


Ubiquitous Science: U-Science, Citizen Science, and the Zooniverse Project

by

Kirk Borne

Department of Computational & Data Sciences

George Mason University

We are facing a huge problem !

The Tsunami

We are facing a huge problem !

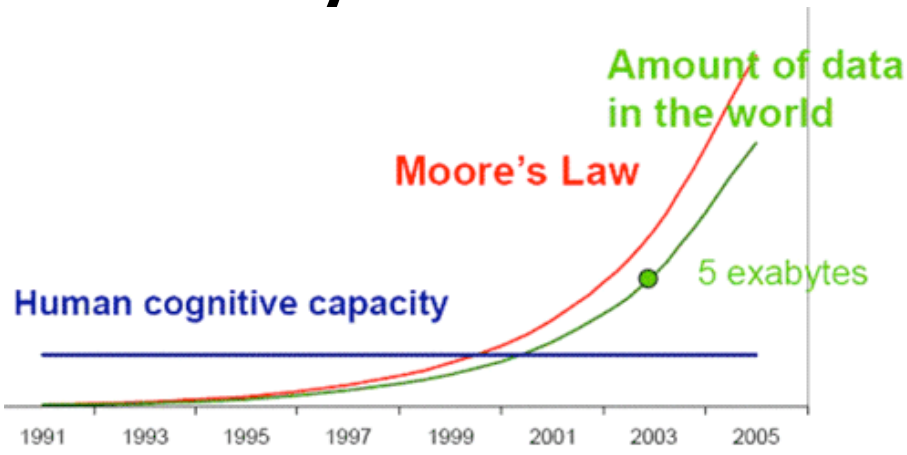
The Data Tsunami

Data Sciences: A National Imperative

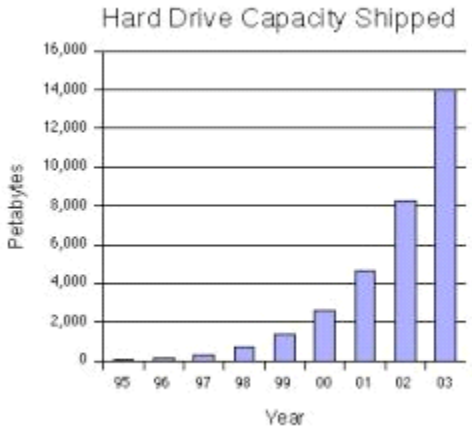
1. National Academies report: *Bits of Power: Issues in Global Access to Scientific Data,* (1997) downloaded from http://www.nap.edu/catalog.php?record_id=5504
2. NSF (National Science Foundation) report: *Knowledge Lost in Information: Research Directions for Digital Libraries,* (2003) downloaded from http://www.sis.pitt.edu/~dlwkshop/report.pdf
3. NSF report: *Cyberinfrastructure for Environmental Research and Education,* (2003) downloaded from http://www.ncar.ucar.edu/cyber/cyberreport.pdf
4. NSB (National Science Board) report: *Long-lived Digital Data Collections: Enabling Research and Education in the 21st Century,* (2005) downloaded from http://www.nsf.gov/nsb/documents/2005/LLDDC_report.pdf
5. NSF report with the Computing Research Association: *Cyberinfrastructure for Education and Learning for the Future: A Vision and Research Agenda,* (2005) downloaded from http://www.cra.org/reports/cyberinfrastructure.pdf
6. NSF Atkins Report: *Revolutionizing Science & Engineering Through Cyberinfrastructure: Report of the NSF Blue-Ribbon Advisory Panel on Cyberinfrastructure,* (2005) downloaded from http://www.nsf.gov/od/oci/reports/atkins.pdf
7. NSF report: *The Role of Academic Libraries in the Digital Data Universe,* (2006) downloaded from http://www.arl.org/bm~doc/digdatarpt.pdf
8. National Research Council, National Academies Press report: *Learning to Think Spatially,* (2006) downloaded from http://www.nap.edu/catalog.php?record_id=11019
9. NSF report: *Cyberinfrastructure Vision for 21st Century Discovery,* (2007) downloaded from http://www.nsf.gov/od/oci/ci_v5.pdf
10. JISC/NSF Workshop report on Data-Driven Science & Repositories, (2007) downloaded from http://www.sis.pitt.edu/~repwkshop/NSF-JISC-report.pdf
11. DOE report: *Visualization and Knowledge Discovery: Report from the DOE/ASCR Workshop on Visual Analysis and Data Exploration at Extreme Scale,* (2007) downloaded from http://www.sc.doe.gov/ascr/ProgramDocuments/Docs/DOE-Visualization-Report-2007.pdf
12. DOE report: *Mathematics for Analysis of Petascale Data Workshop Report,* (2008) downloaded from http://www.sc.doe.gov/ascr/ProgramDocuments/Docs/PetascaleDataWorkshopReport.pdf
13. NSTC Interagency Working Group on Digital Data report: *Harnessing the Power of Digital Data for Science and Society,* (2009) downloaded from http://www.nitrd.gov/about/Harnessing_Power_Web.pdf
14. National Academies report: *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age,* (2009) downloaded from http://www.nap.edu/catalog.php?record_id=12615

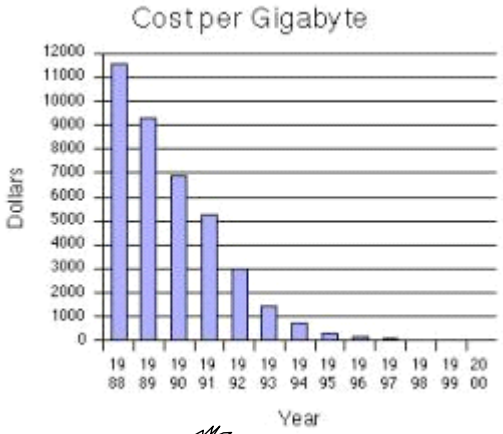
Data Doubles Every Year

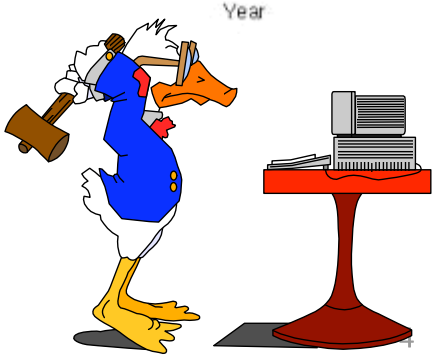
- Computing power doubles every 18 months (Moore's Law) ...
  - **100x in 10 years**
- I/O bandwidth increases ~10% / year
  - **<3x in 10 years**.
- Data doubles every year ...
  - **1000x in 10 years**, and 1,000,000x in 20 yrs.
  - NCSA Example:
    - First 19 years: 1 PB
    - Year 20 (2007): 2 PB
    - Year 21 (2008): 4 PB
    - By 2020 : ~20 Exabytes ?
  - *In the Year 2525*: 10^{156} PB ?????
- As our data volumes grow, especially in the sciences (where scientific funding for research barely grows at all), we will fall farther and farther behind in our ability to analyze, assimilate, and extract knowledge from our data collections ... unless we develop and apply exponentially more powerful algorithms and methods.



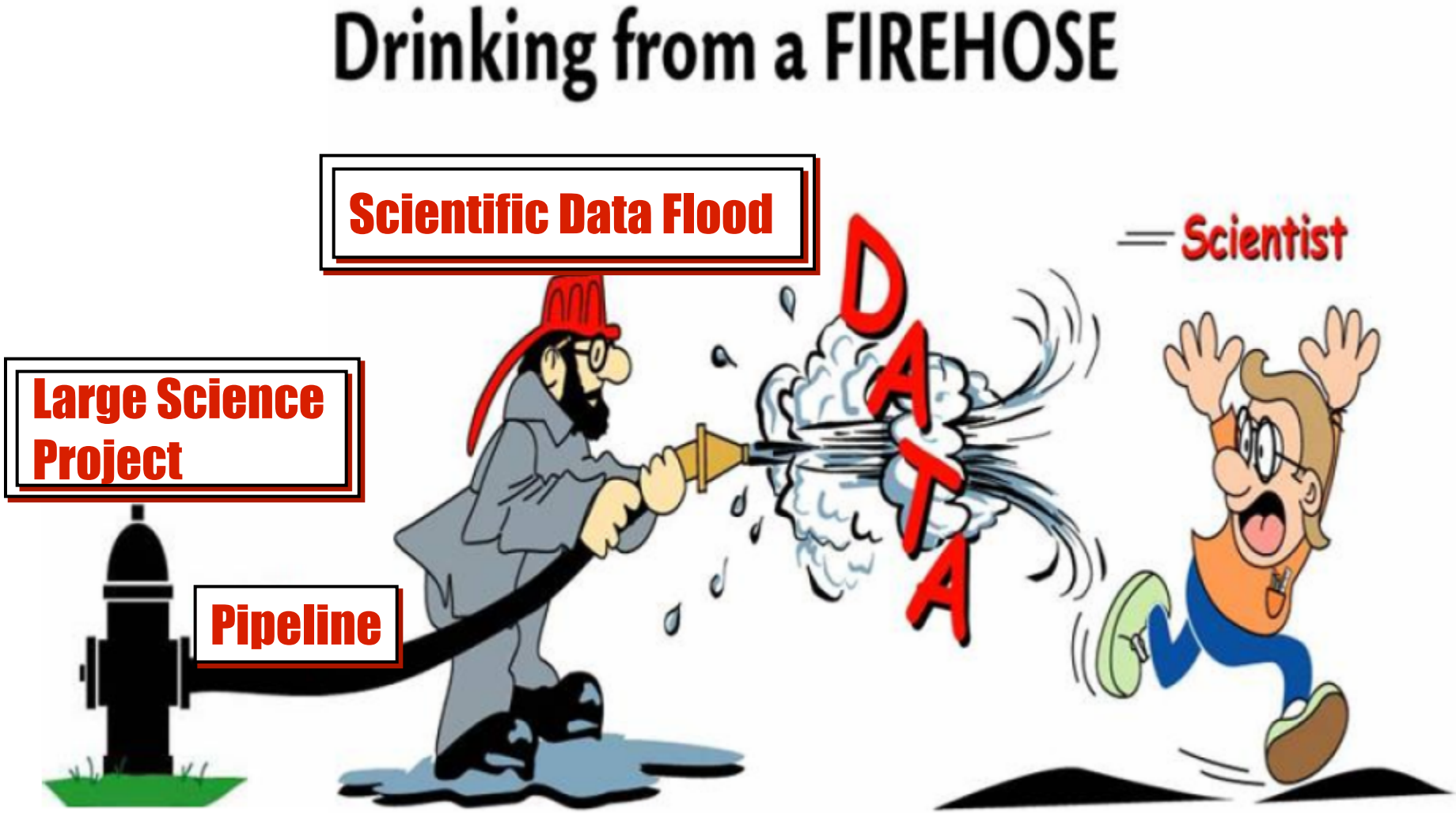
(Illustration adapted from a slide by J. Heer, PARC User Interface Research Group)







The Scientific Data Flood



Characteristics of the Data Flood in all of the Sciences

- **Large** quantities of data are acquired.
- But … what do we mean by "**large**"?
  - Gigabytes? Terabytes? Petabytes? Exabytes?
  - The meaning of "large" is domain-specific and resource-dependent (data storage, I/O bandwidth, computation cycles, communication costs)
  - I say … we all are dealing with our own "**tonnabytes**"
- There are 3 dimensions to the data flood problem:
  1. **Volume** (*tonnabytes challenge*)
  2. **Complexity** (*curse of dimensionality*)
  3. **Multiple modalities** (*data fusion and semantic integration challenge*)
- **Challenges**: fusion, integration, knowledge discovery
- Therefore, we need something better to cope with the data tsunami … **the Tonnabytes** !

Outline

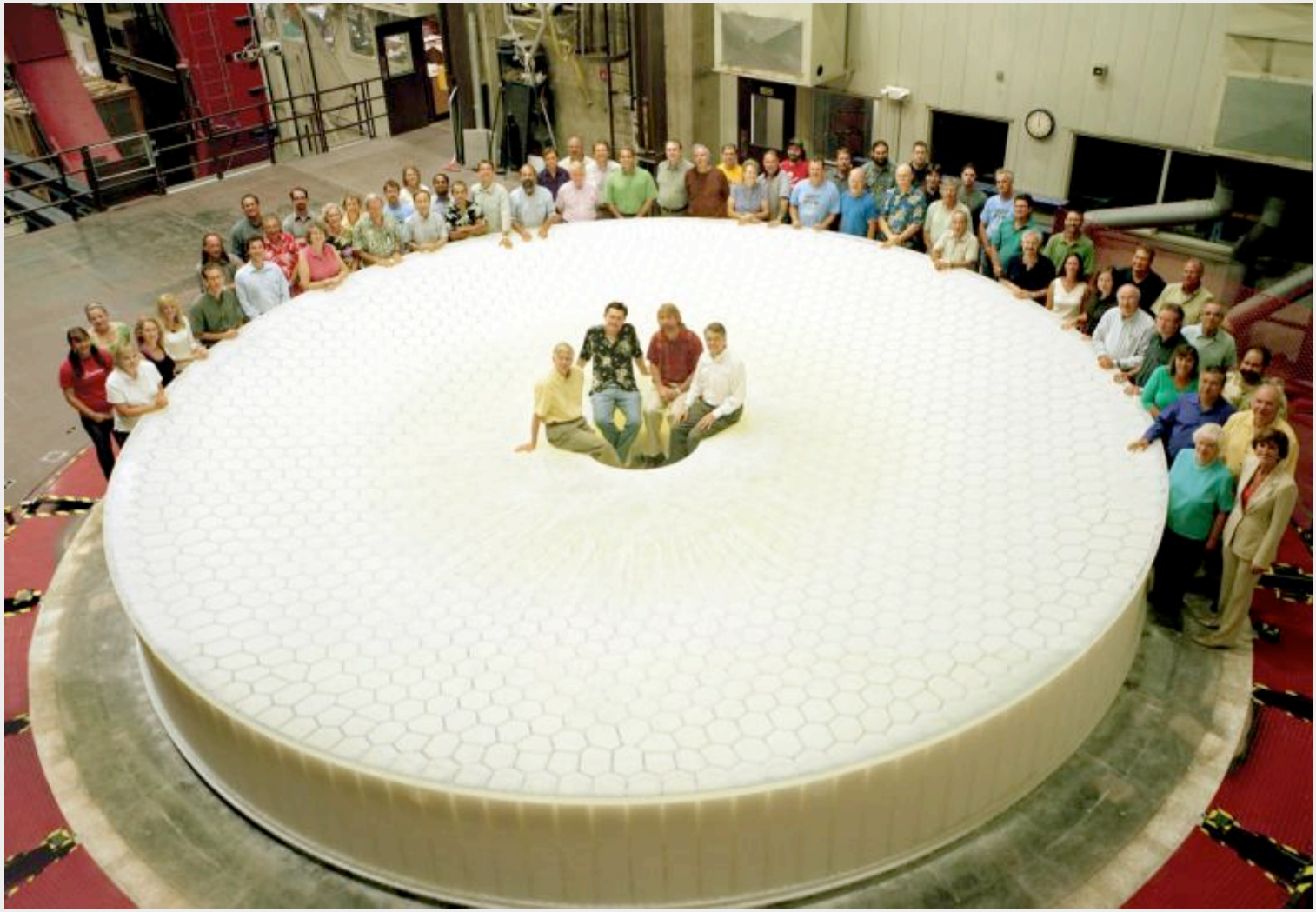
- Astronomy example: the LSST Project
  - Classification of millions of real-time events
  - Classification of billions of galaxies
- Human Computation
- U-Science
- The Zooniverse Project

Outline

- Astronomy example: the LSST Project
  - Classification of millions of real-time events
  - Classification of billions of galaxies
- Human Computation
- U-Science
- The Zooniverse Project

Example project:
The Large Synoptic Survey Telescope

http://www.lsst.org/

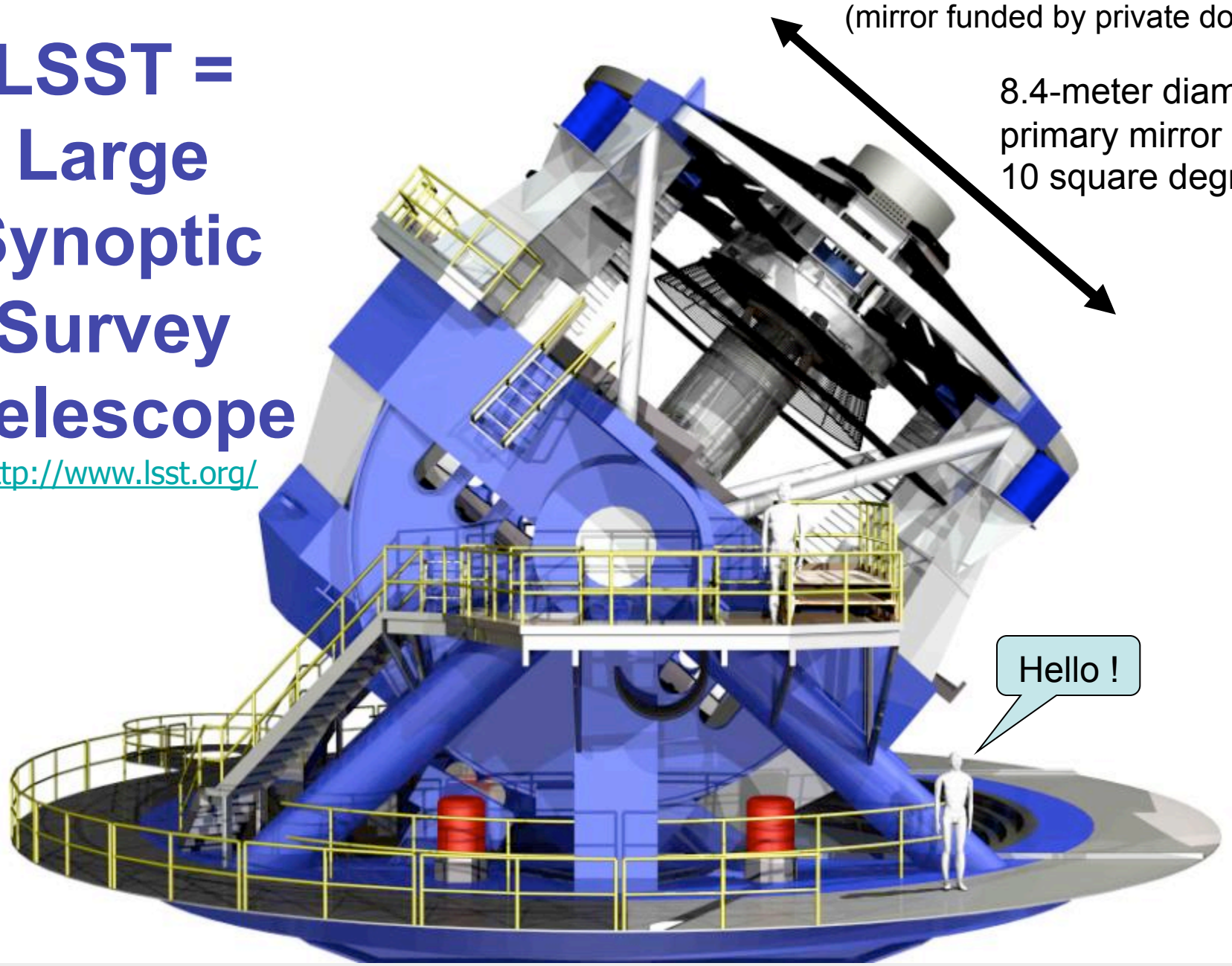


This is the real LSST mirror (August 2008), which will be figured and polished over the next 6 years

(notice the cool guy in the middle of the back row in the gold LSU tigers shirt)

LSST = Large Synoptic Survey Telescope

http://www.lsst.org/



(design, construction, and operations of telescope, observatory, and data system: NSF) (camera: DOE)

LSST Key Science Drivers: Mapping the Universe

– **Solar System Map** (moving objects, NEOs, asteroids: census & tracking)
– **Nature of Dark Energy** (distant supernovae, weak lensing, cosmology)
– **Optical transients** (of all kinds, with alert notifications within 60 seconds)
– **Galactic Structure** (proper motions, stellar populations, star streams)

South America

Chile

Region de Coquimbo

Summit of Cerro Pachon -

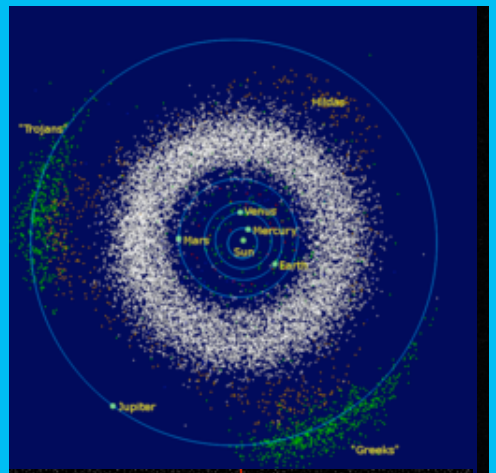
Model of LSST Observatory

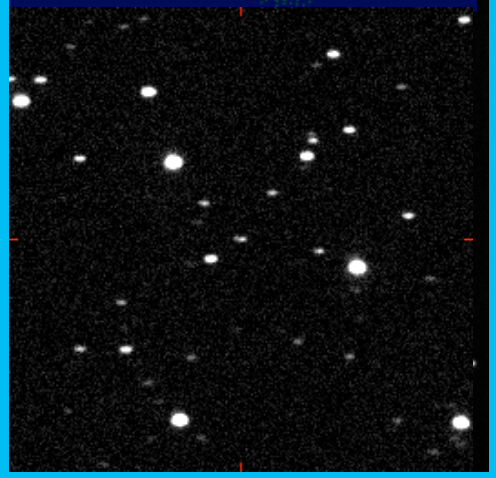
LSST in time and space:

– When? 2016-2026
– Where? Cerro Pachon, Chile

Dynamic Astronomy: The Universe is not static !

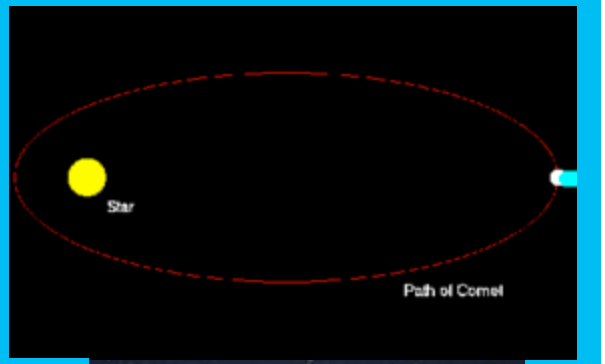
Asteroids





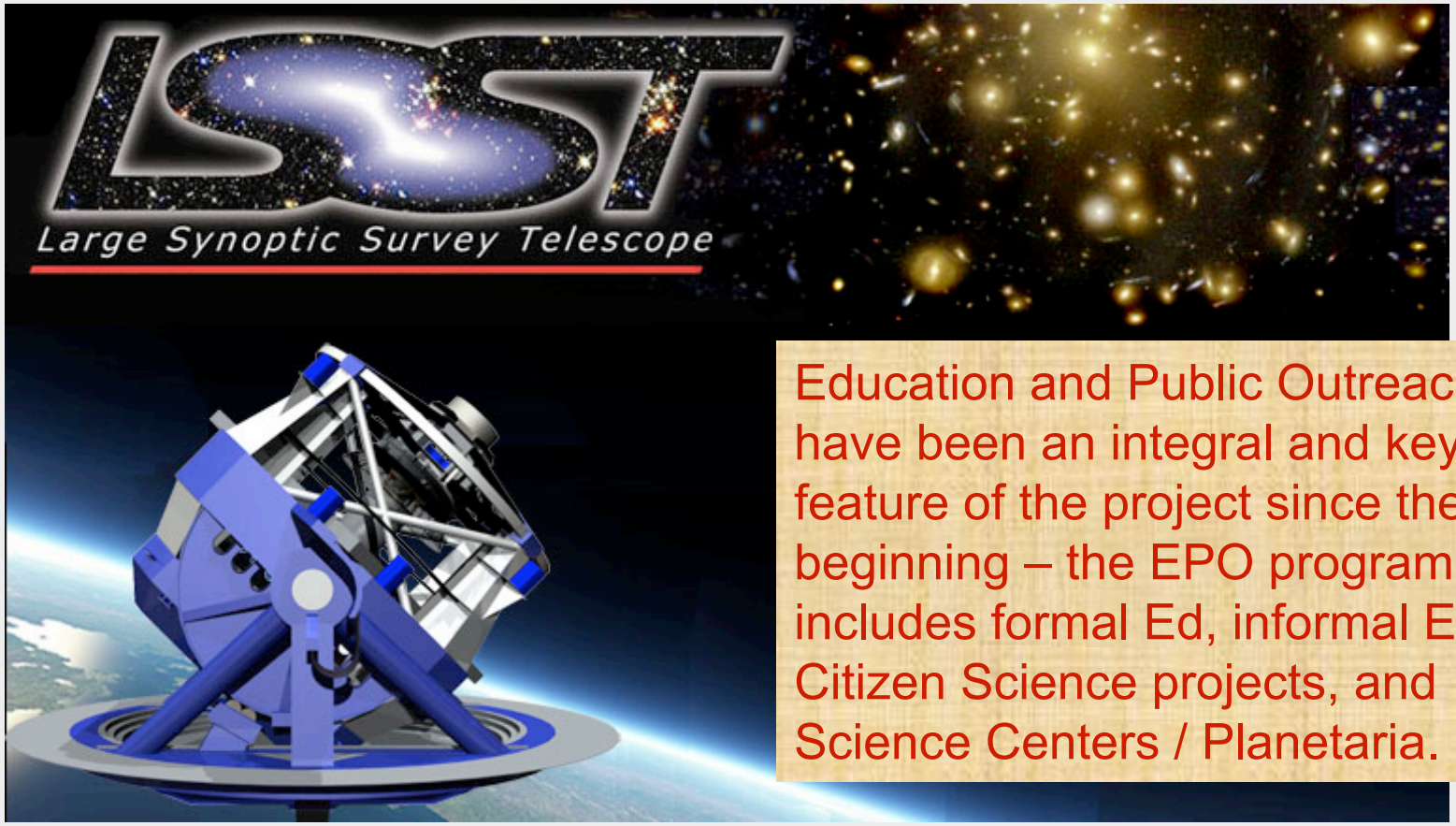
Supernovae

Comets



Observing Strategy: One pair of images every 40 seconds for each spot on the sky, then continue across the sky continuously every night for 10 years (2016-2026), with time domain sampling in log(time) intervals (to capture dynamic range of transients).

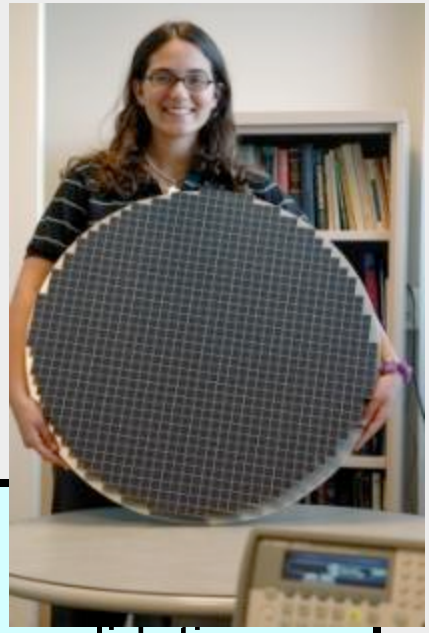
- **LSST (Large Synoptic Survey Telescope):**
  - Ten-year time series imaging of the night sky – mapping the Universe !
  - **100,000 events each night** – *anything that goes bump in the night !*
  - ***Cosmic Cinematography! The New Sky! @ http://www.lsst.org/***



Education and Public Outreach have been an integral and key feature of the project since the beginning – the EPO program includes formal Ed, informal Ed, Citizen Science projects, and Science Centers / Planetaria.

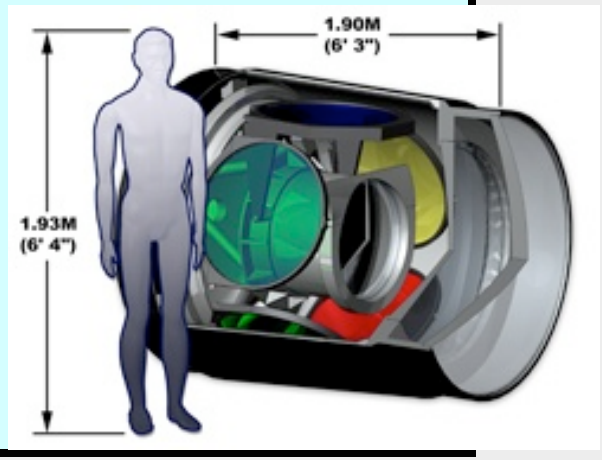
The LSST focal plane array

Camera Specs: **(pending funding from the DOE)**
201 CCDs @ 4096x4096 pixels each!
= 3 Gigapixels = 6 GB per image, covering 10 sq.degrees
= ~3000 times the area of one Hubble Telescope image



LSST Data Challenges

- Obtain one 6-GB sky image in 15 seconds
- Process that image in 5 seconds
- Obtain & process another co-located image for science validation within 20^{s} (= 15-second exposure + 5-second processing & slew)
- Process the 100 million sources in each image pair, catalog all sources, and generate worldwide alerts within 60 seconds (e.g., incoming killer asteroid)
- Generate 100,000 alerts per night (VOEvent messages)
- Obtain 2000 images per night
- Produce ~30 Terabytes per night
- Move the data from South America to US daily
- Repeat this every day for 10 years (2016-2026)
- Provide rapid DB access to worldwide community:
  - **100-200 Petabyte image archive**
  - **10-20 Petabyte database catalog**

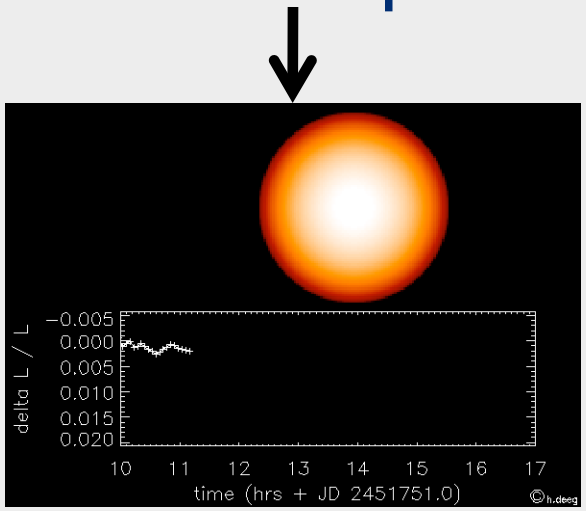


Two challenge problems with LSST:

1. Classification of real-time event stream of 100,000 events every night for 10 years
2. Classification of 50 billion astronomical objects in 100 Petabytes of images, including over 20 billion galaxies

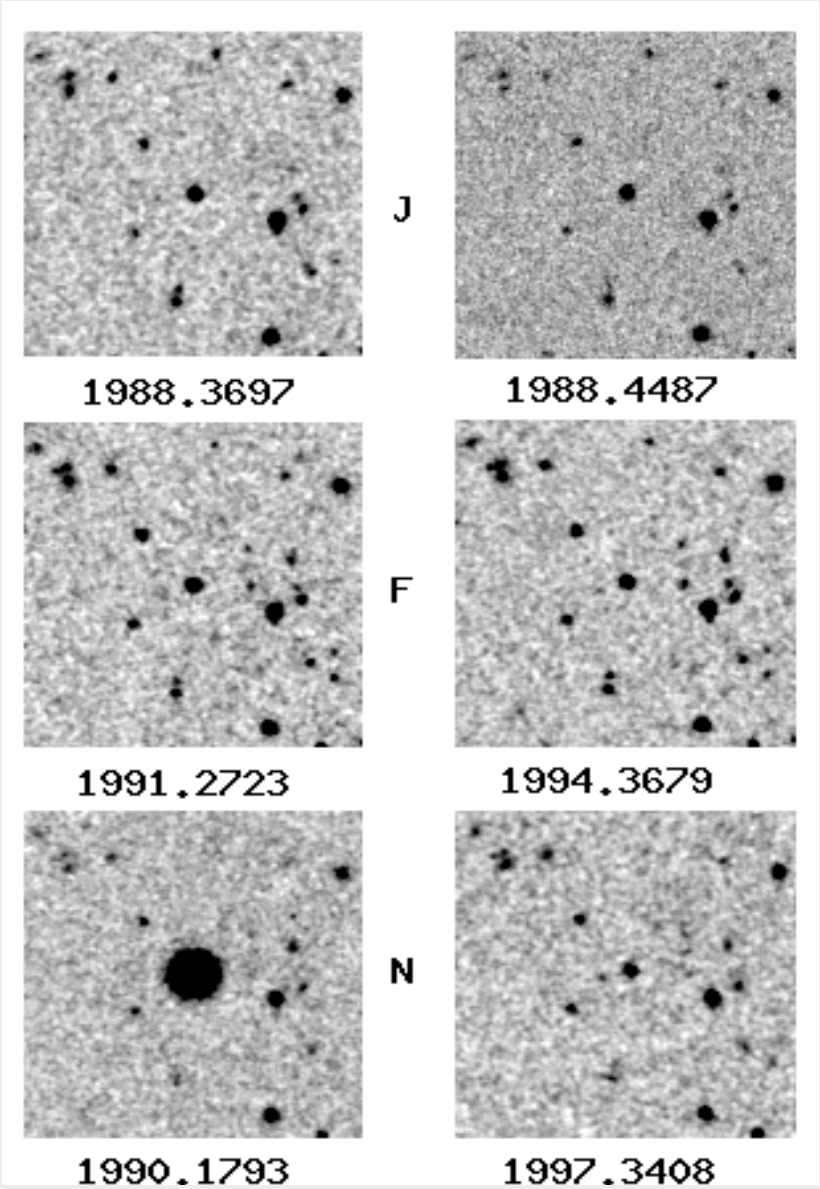
What is an Event?

- Anything that changes (motion or brightness)
- Variable stars of all kinds
- Optical transients: e.g., extra-solar planets
- Supernova
- Gamma-ray burst
- New comet
- New asteroid
- Incoming Killer Asteroid
- Anything that goes bump in the night

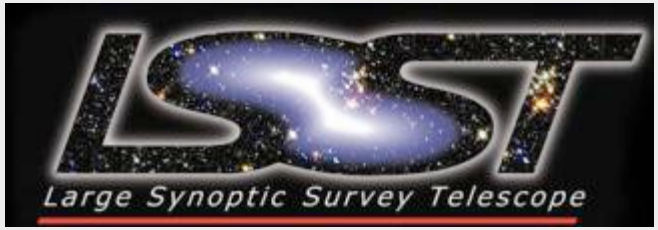


Here is one type of event ***

- **Optical Transient:** here today, gone tomorrow
- It is a normal dwarf star, similar to our sun, except …
- **it increased in brightness by 300x in one night** …
- and then returned to normal.



***Courtesy: Caltech / Palomar Quest Survey



The LSST will represent a 100,000-fold increase in the VOEvent network traffic and in the real-time classification demands:

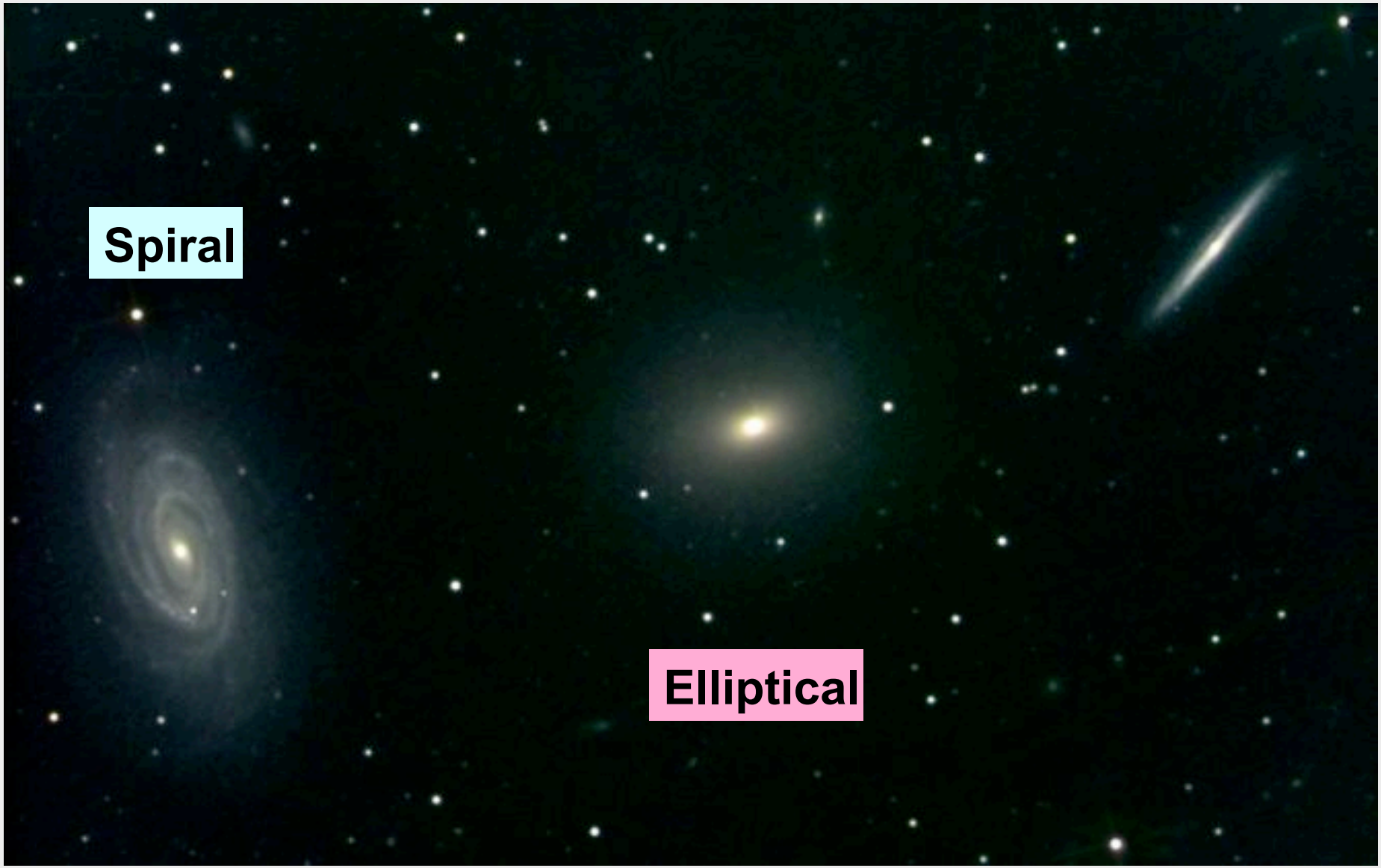
from data to knowledge!

from sensors to sense!

Two challenge problems with LSST:

1. Classification of real-time event stream of 100,000 events every night for 10 years
2. Classification of 50 billion astronomical objects in 100 Petabytes of images, including over 20 billion galaxies

There are 2 main types of galaxies: Spiral & Elliptical (plus there are some peculiar & irregular galaxies)

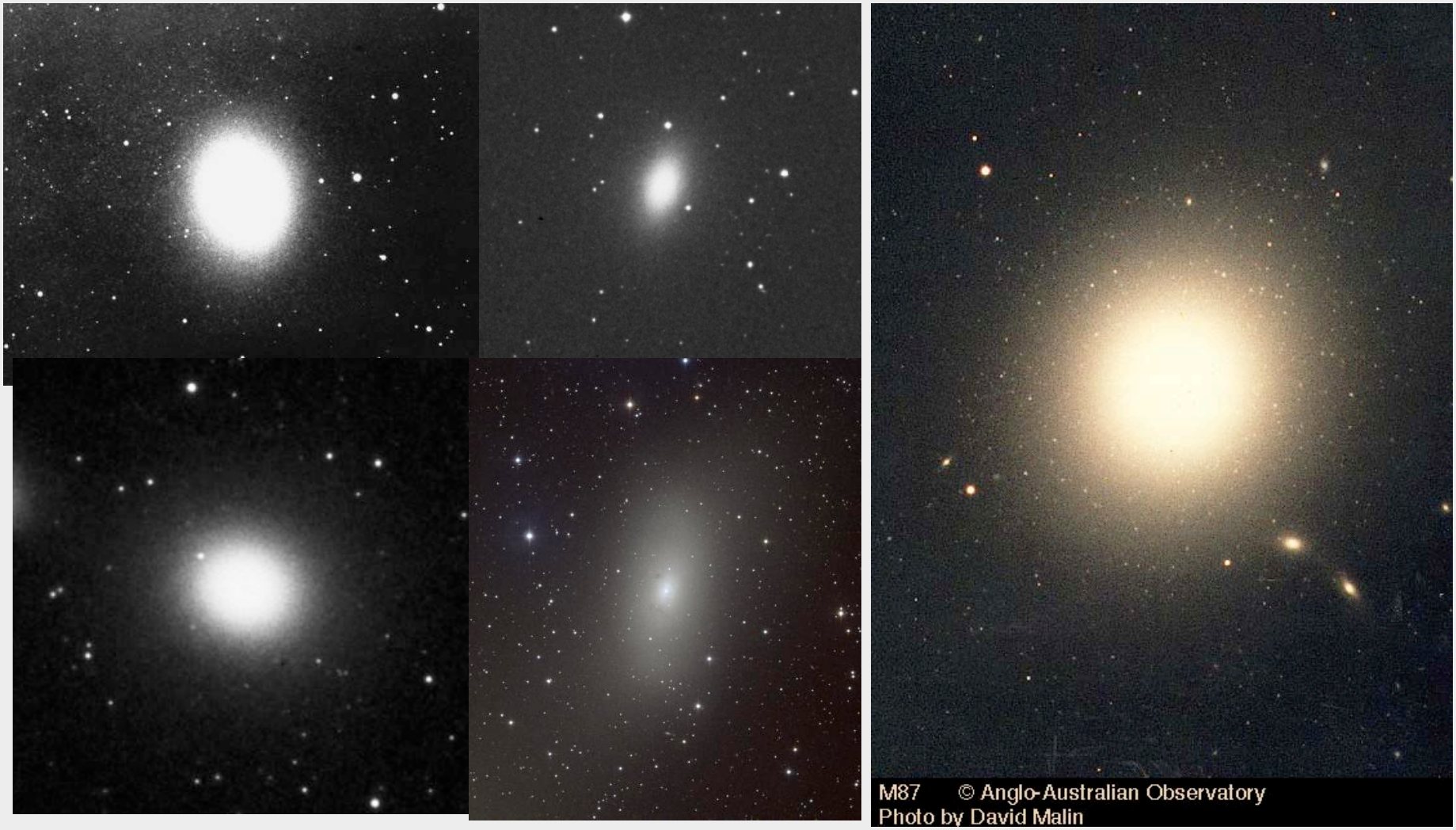


Gallery of Elliptical Galaxies

M32

M59

M87



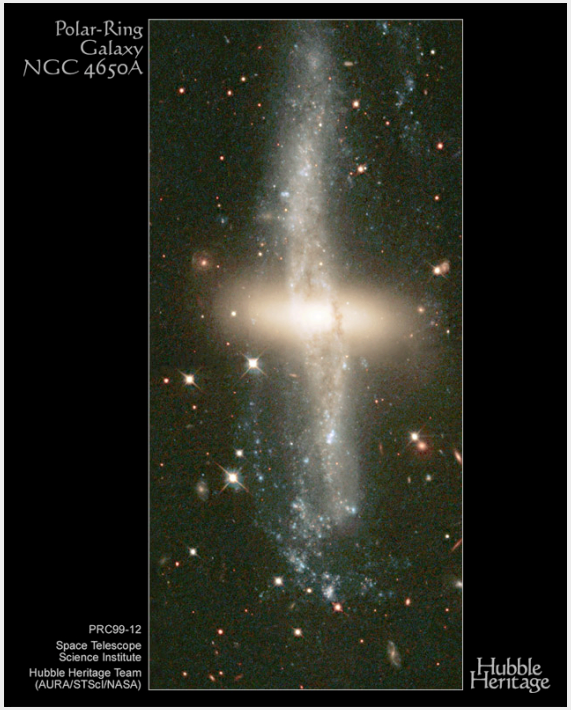
M105

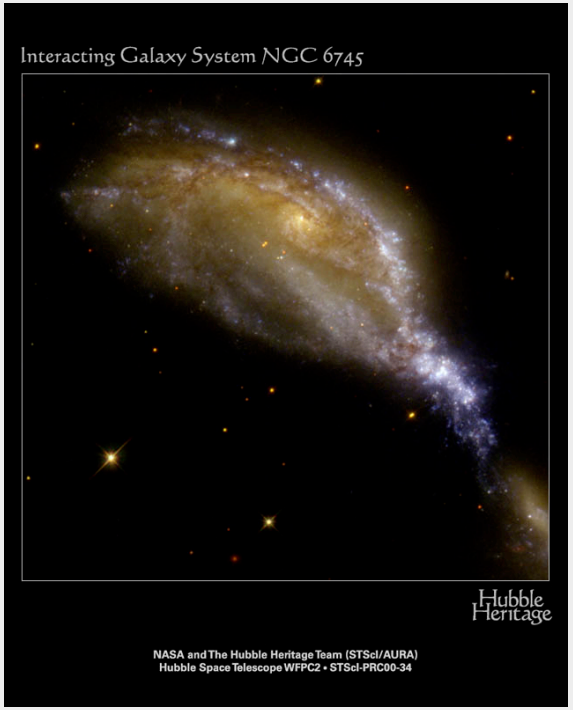
M110=NGC205

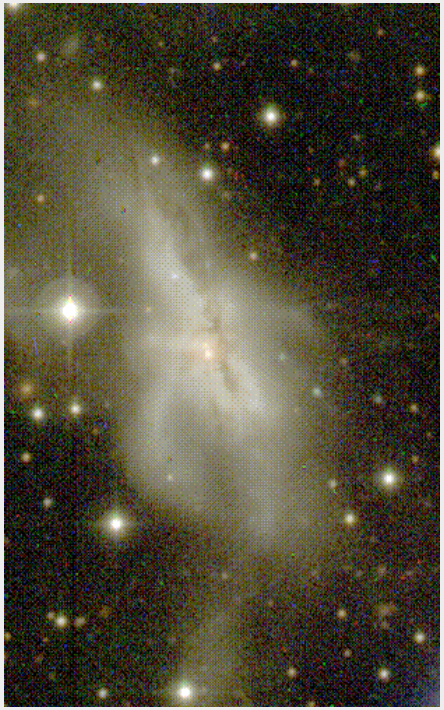
Gallery of Face-on Spiral Galaxies: studying their properties indicates that our Milky Way is a Spiral

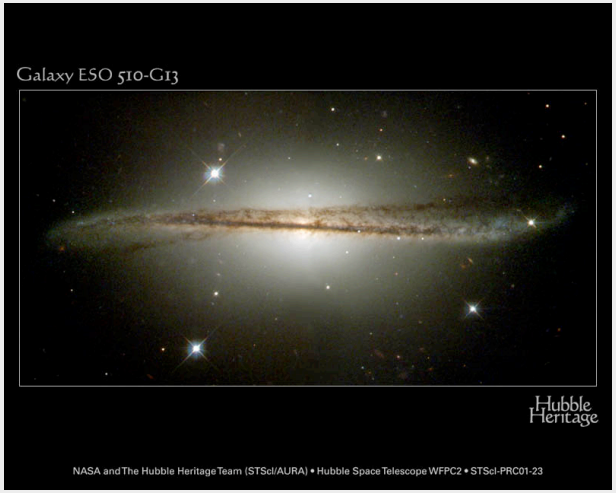
Gallery of Edge-on Spiral Galaxies

There are lots of Peculiar Galaxies also !



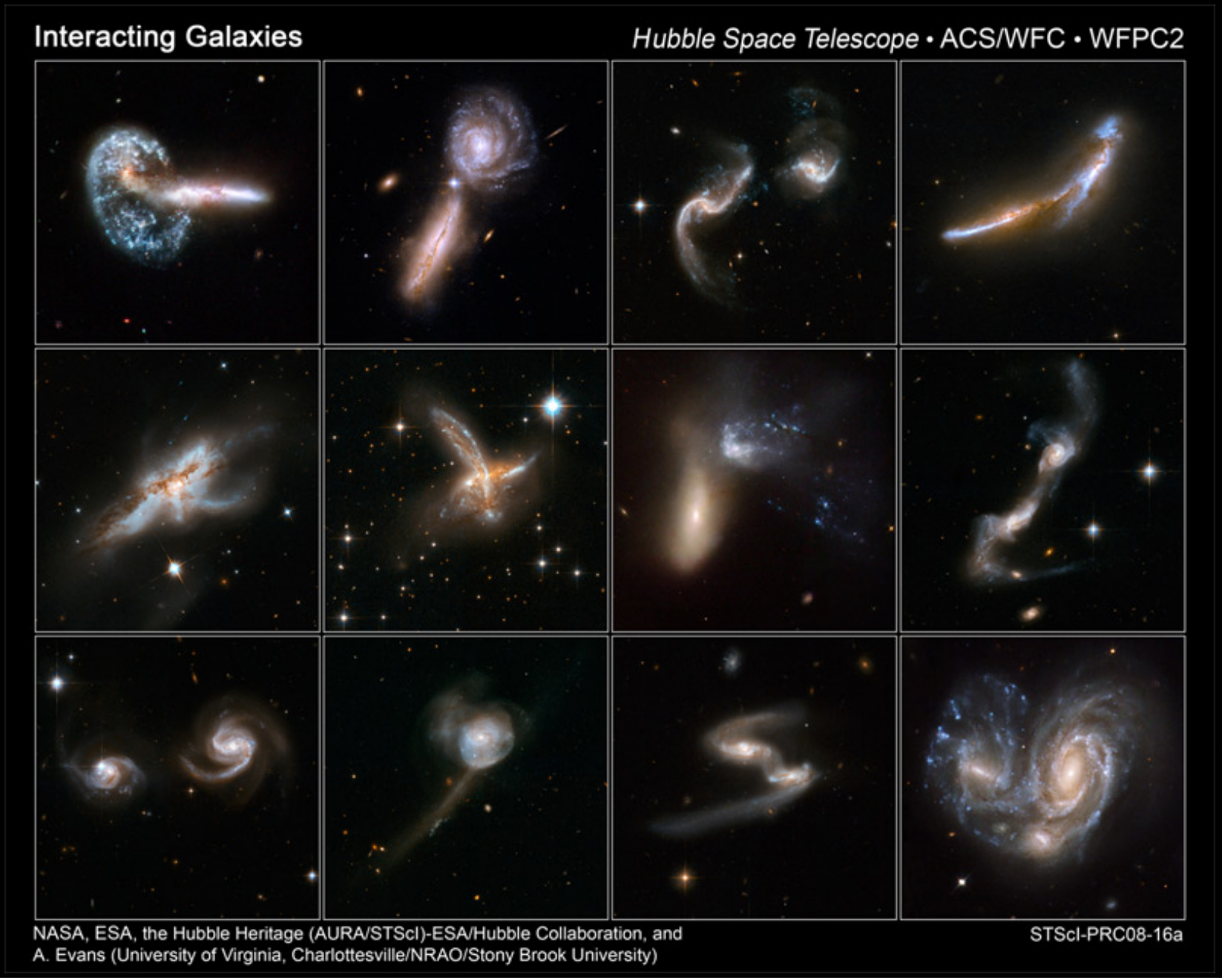




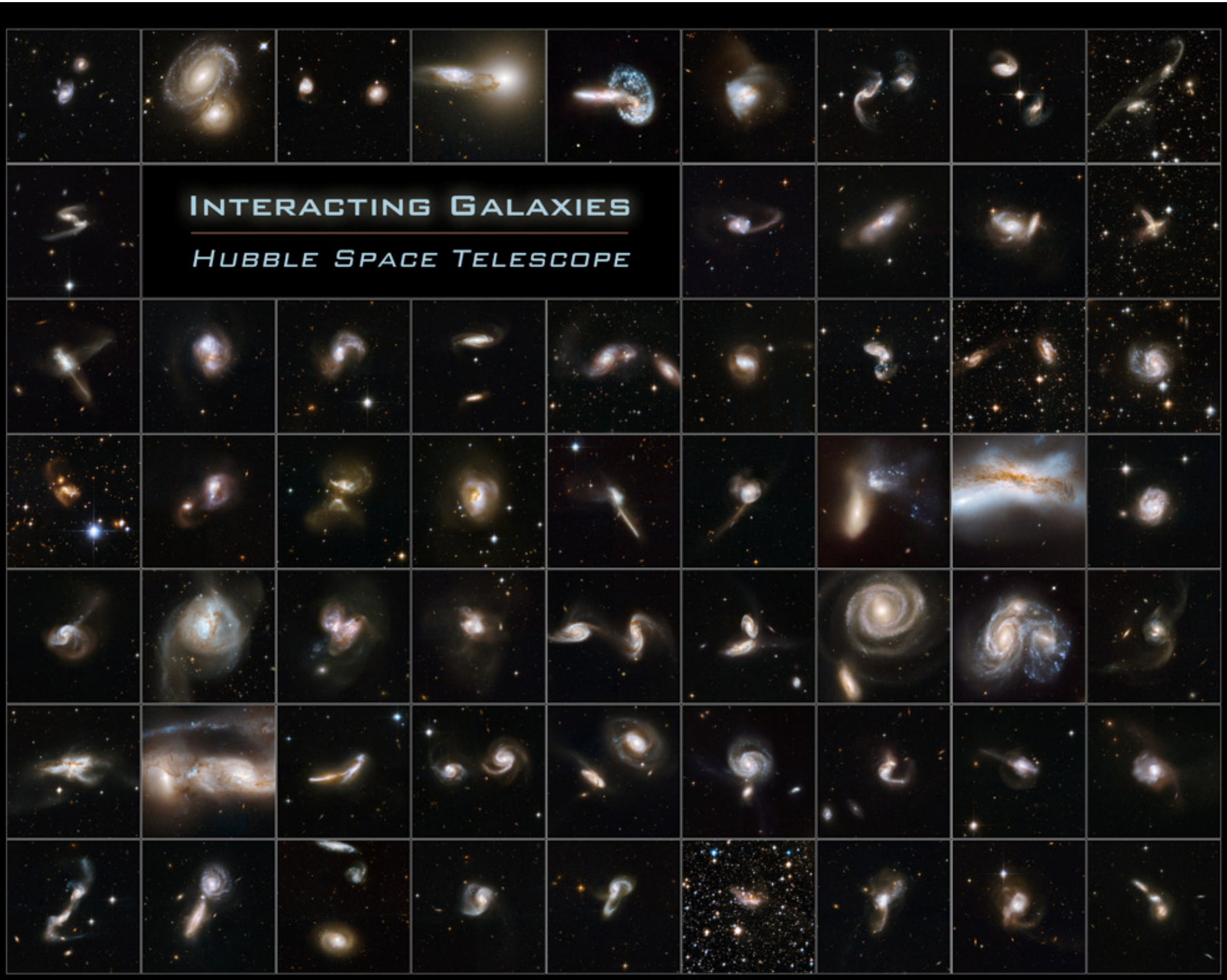


Galaxies Gone Wild !

Colliding and Merging Galaxies = Interacting Galaxies



Colliding and Merging Galaxies = Interacting Galaxies


Interacting Galaxies
Hubble Space Telescope

Merging/Colliding Galaxies are the building blocks of the Universe: 1+1=1

More gorgeous Colliding Galaxies !

NGC 6050

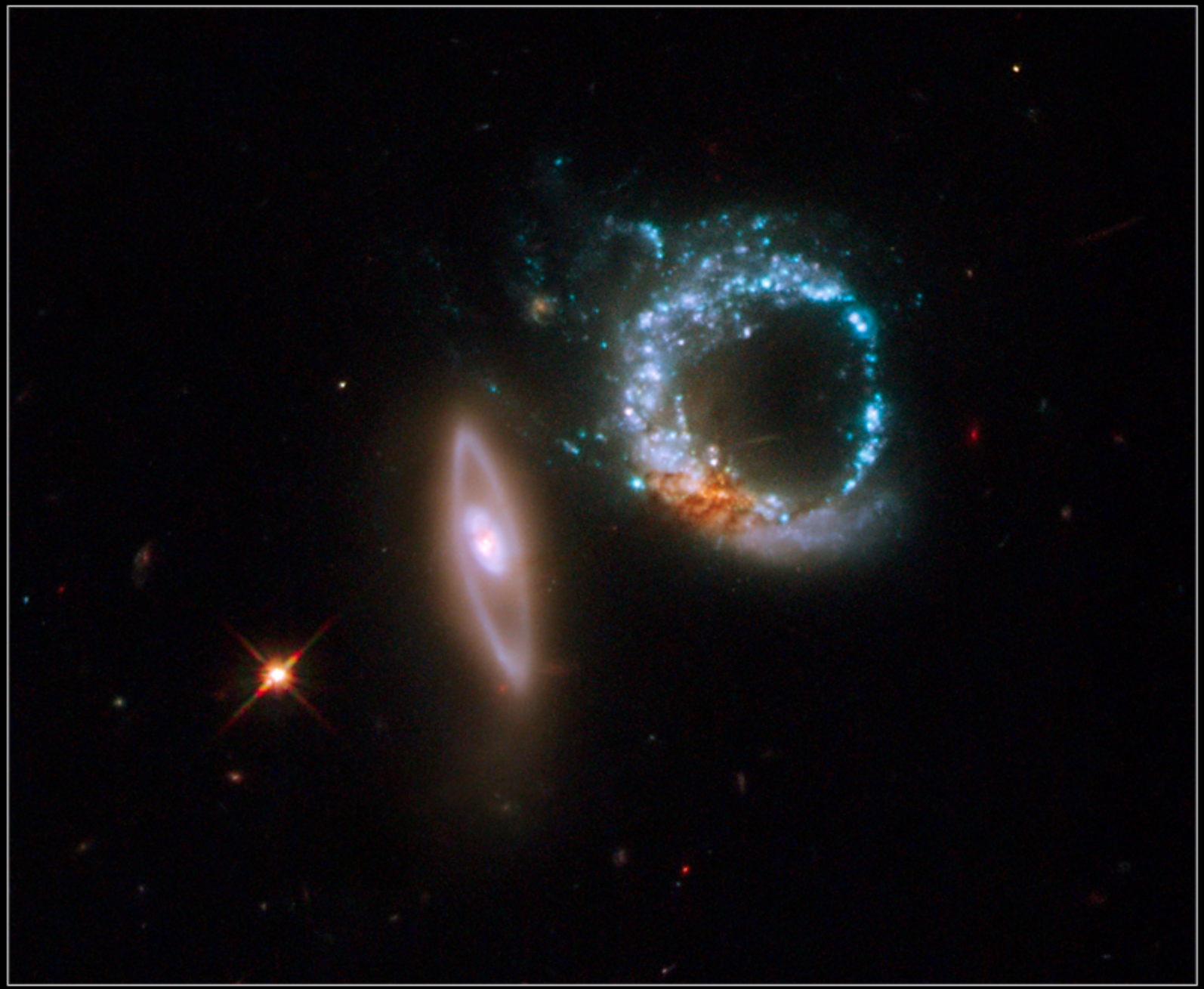
Arp 148

Hubble Heritage

NASA, ESA, and The Hubble Heritage (AURA/STScI)-ESA/Hubble Collaboration • *Hubble Space Telescope* WFPC2

Interacting Galaxies Arp 147

Hubble Space Telescope • WFPC2



NASA, ESA, and M. Livio (STScI)

STScI-PRC08-37

The Universe is full of galaxies

Astronomers have collected images of hundreds of millions of galaxies, but we have analyzed maybe only 10% of these!

In the next 10-20 years, we will have images of tens of **billions** of new galaxies

How will we identify and classify all of these spirals, ellipticals, and mergers?

Spirals, ellipticals, and mergers? Oh my!

The LSST Data Challenges

100 PB image archive

2 TB of data per hour

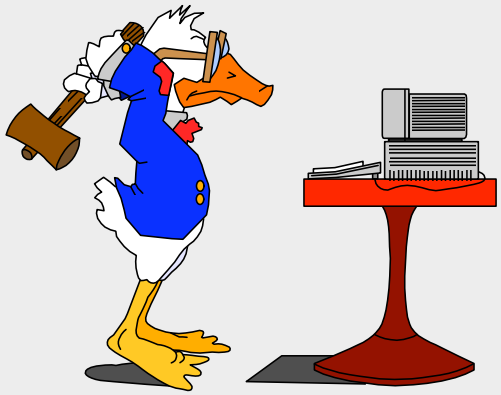
100,000 events every night

50 billion object database

20 PB science catalog

The LSST Data Challenges

How will we respond ?



We need something better …

We need something better, Jim !

Outline

- Astronomy example: the LSST Project
  - Classification of millions of real-time events
  - Classification of billions of galaxies
- Human Computation
- U-Science
- The Zooniverse Project

We need computers …
but not the usual kind !

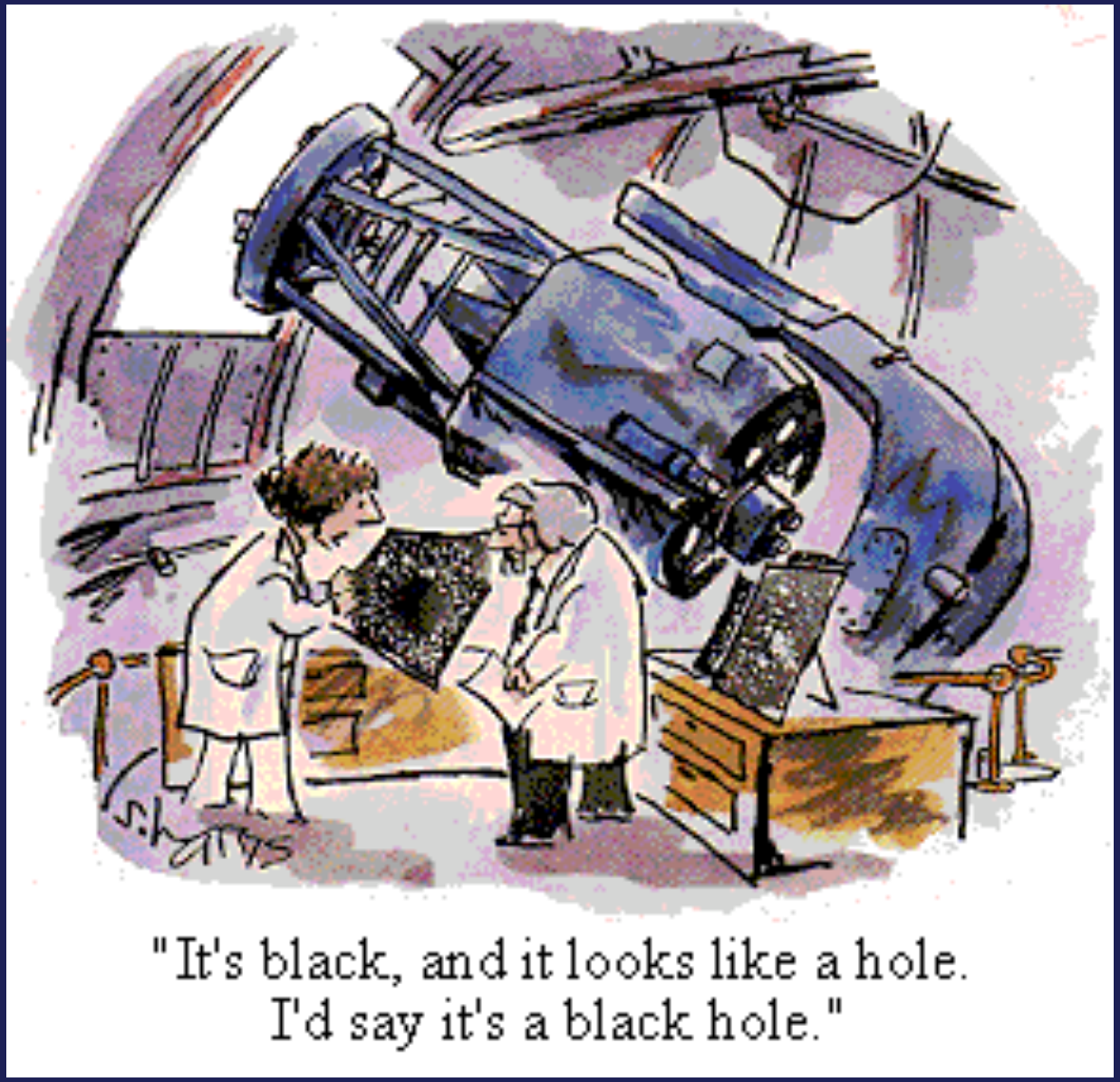
We need the classical kind (which pre-dates computing devices)

Modes of Computing

- **Numerical Computation**
  - Fast, efficient
  - Processing power is rapidly increasing
  - Model-dependent, subjective, only as good as your best hypothesis
- **Computational Intelligence**
  - Data-driven, objective (machine learning)
  - Often relies on human-generated training data
  - Often generated by a single investigator
  - Primitive algorithms
  - Not as good as humans on most tasks
- **Human Computation (computational thinking)**
  - Data-driven, objective (human cognition)
  - Creates training sets, Cross-checks machine results
  - Excellent at finding patterns, image classification
  - Capable of classifying anomalies that machines don't understand
  - Slow at numerical processing, low bandwidth, easily distracted

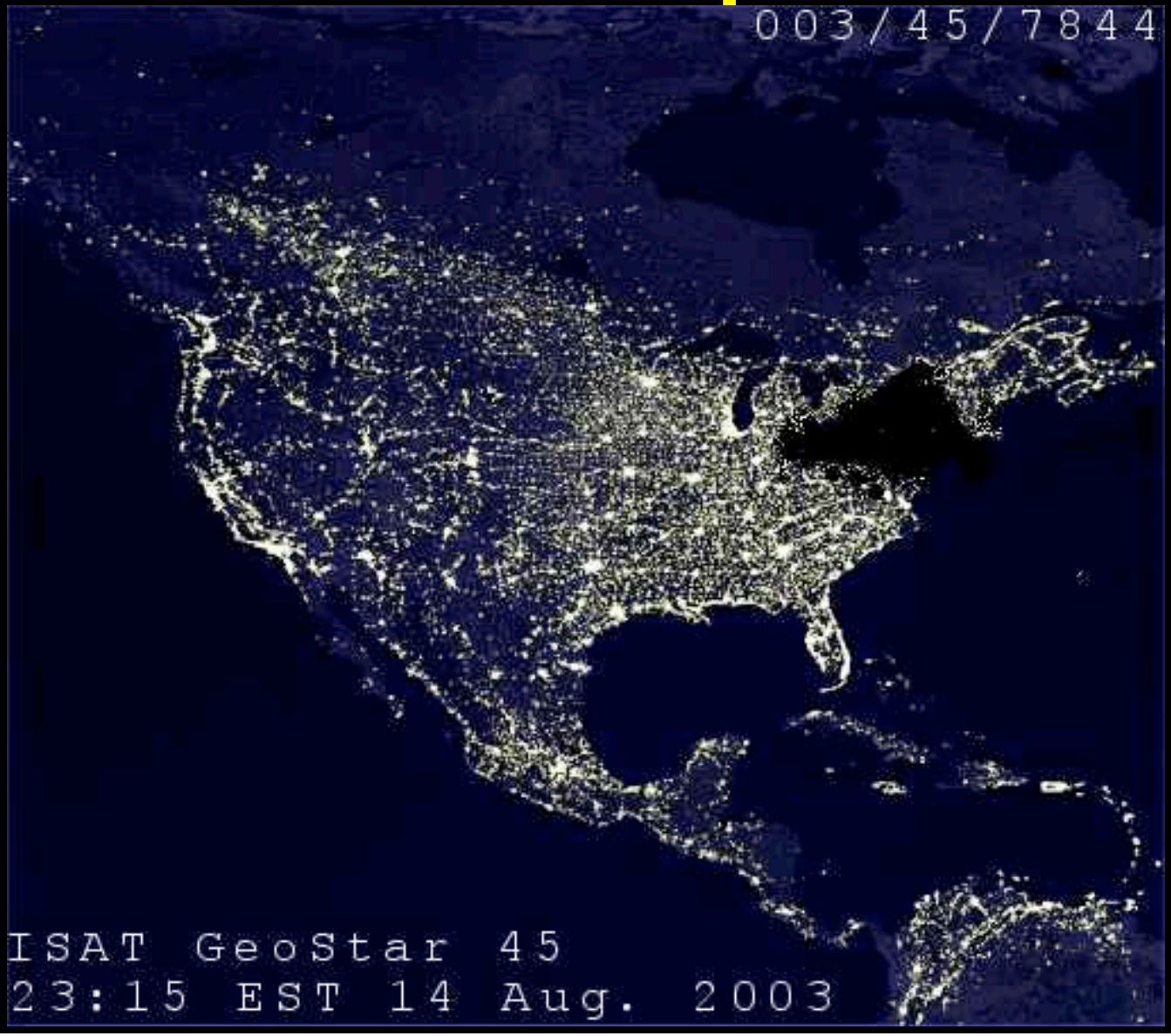
It takes a human to interpret a complex image

It takes a human to interpret a complex image ... usually ...



"It's black, and it looks like a hole.
I'd say it's a black hole."

What words would you use to describe this picture?



Enter ....

U-Science

Outline

- Astronomy example: the LSST Project
  - Classification of millions of real-time events
  - Classification of billions of galaxies
- Human Computation
- U-Science
- The Zooniverse Project

U-Science

- The emergence of Citizen Science
- Science@home
- Science 2.0
- Anybody can participate in the science discovery process
- Anyone can annotate, tag, and label scientific results: scientists, students, and citizen scientists

What is U-Science?

- User-centered, User-driven science
- Ubiquitous
- Universal
- Unleashed
- Untethered [http://tw.rpi.edu]
- “You”-centric:
  - Think ... Social Networks ...
    - Facebook, Myspace, Youtube
    - Blogs, Wikis, Collaboratories
    - Del.icio.us, flickr.com
- Semantic e-Science:
    - BioDAS, AstroDAS, Wikiproteins, tags, annotations

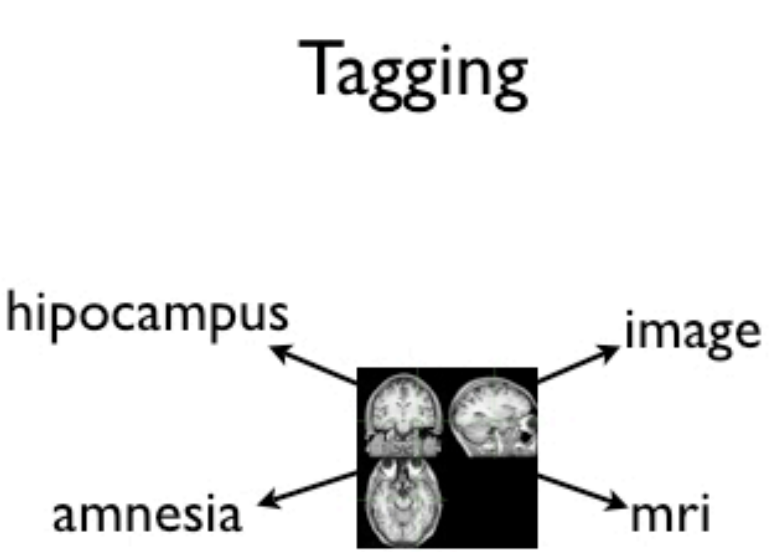
U-Science: it is all about "U"

- If classic web is Web 1.0, then the future Semantic Web is Web 3.0.
- Between these two is the emerging human-friendly, human-powered, and social network-oriented Web 2.0:
  - Wikis and Mashups (e.g., Yahoo Pipes, Microsoft Popfly)
  - Tagging, Annotation, Folksonomies, Microformats, Tag Clouds
  - For example: http://www.flickr.com/ or http://del.icio.us/
  - Human Computation: The ESP game @ http://www.espgame.org/
    - created by Luis Von Ahn (MacArthur genius grant winner)
  - Science 2.0: http://openwetware.org/wiki/Science_2.0/Brainstorming
  - Biology Distributed Annotation System: http://biodas.org/
  - Heliophysics Knowledgebase: http://www.lmsal.com/helio-informatics/hpkb/
  - Wikiproteins: http://www.wikiprofessional.info/
  - Entity Describer: http://www.entitydescriber.org/about.html
  - AstroDAS: coming soon (hopefully!)

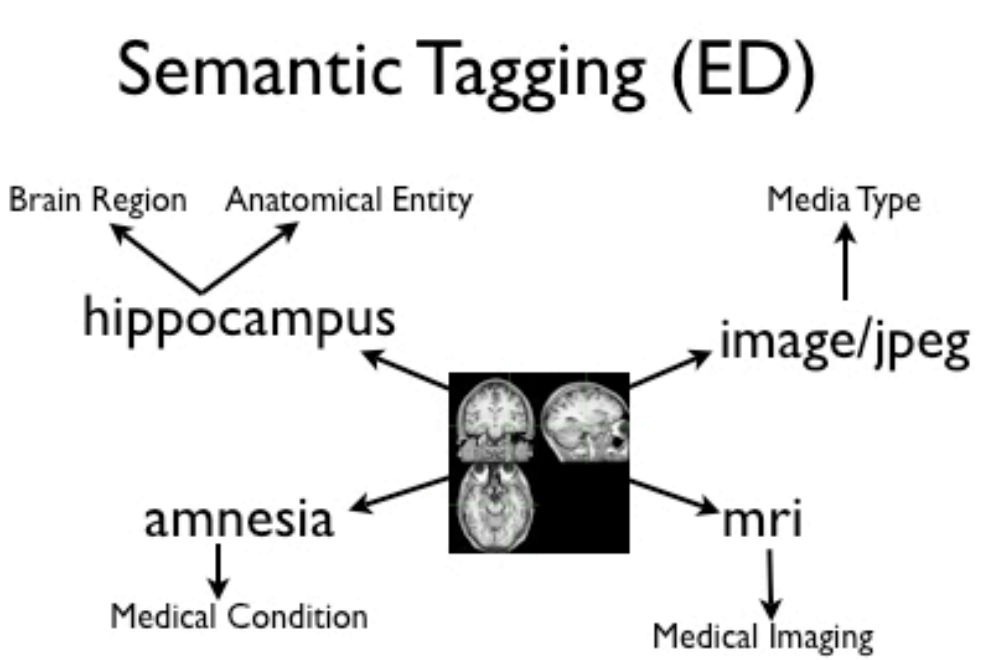
Entity Describer – example

http://www.entitydescriber.org/about.html

- Before semantic tagging:



- After semantic tagging:

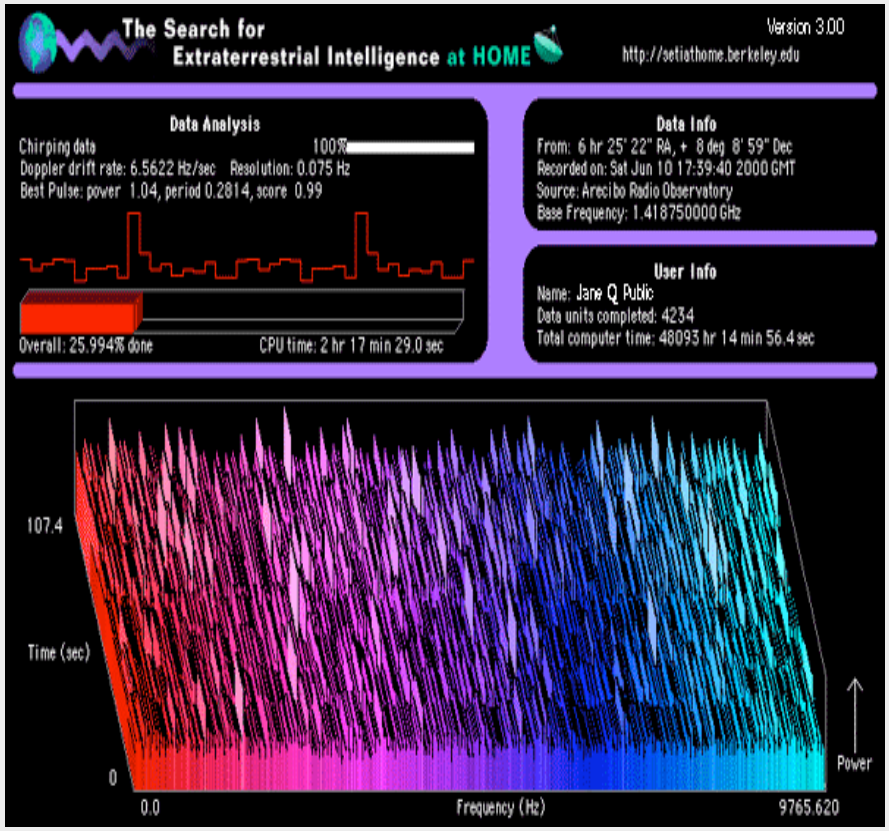


U-Science Examples

- Science@home
- Citizen Science

Science@home examples

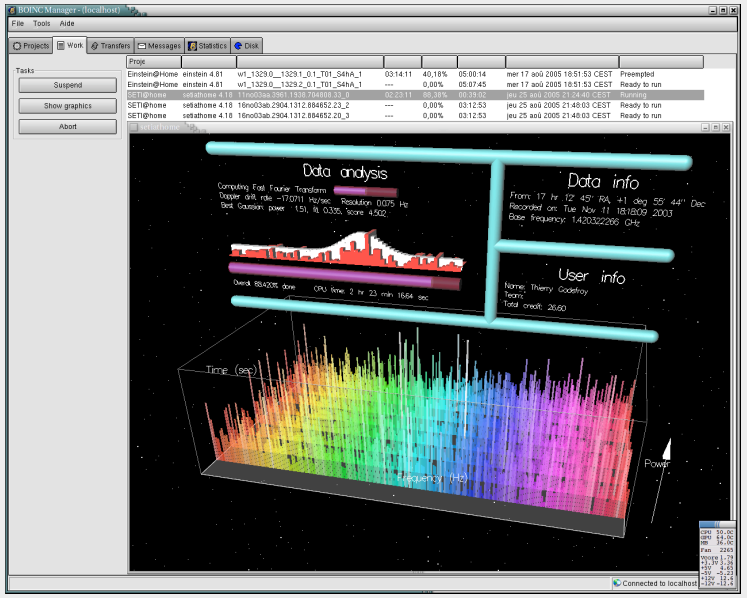
- SETI@home
- Milkyway@home
- Cosmology@home
- Folding@home
- Docking@home
- CELS@home
- Einstein@home
- LHC@home
- Climate@Home **(developed @ GSFC)**
- etc.

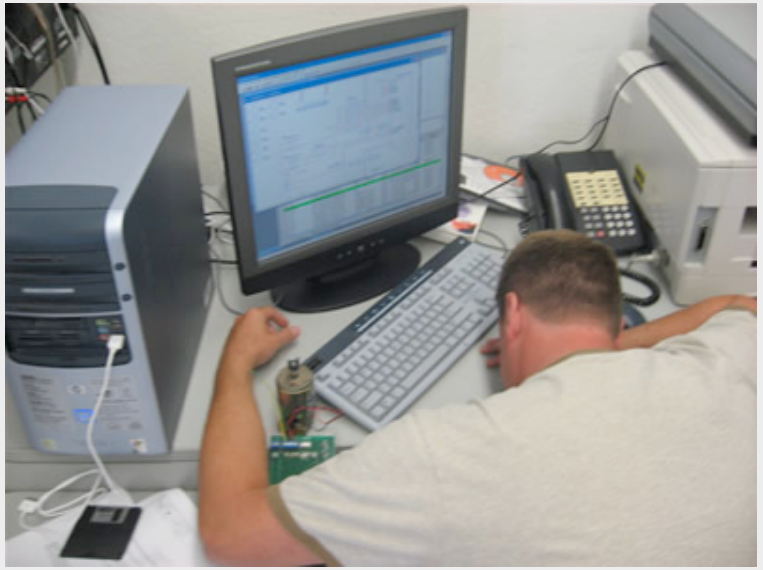


Check out http://boinc.berkeley.edu/projects.php

But those are mostly passive

- Most Science@home projects use your idle computer time, consequently they are immensely powerful engines of scientific productivity.
- Most of those projects do not require you to do anything more than download the screensaver software for your computer.
- The program wakes up when you and your computer go to sleep.





What if you want to do more?

- What if you want to be an active participant in the science discovery process?
- Well, you can.
- You can be actively involved with the scientific discovery process and the large science experiments.
- **Citizen Science !** http://www.citizensci.com/

Citizen Science

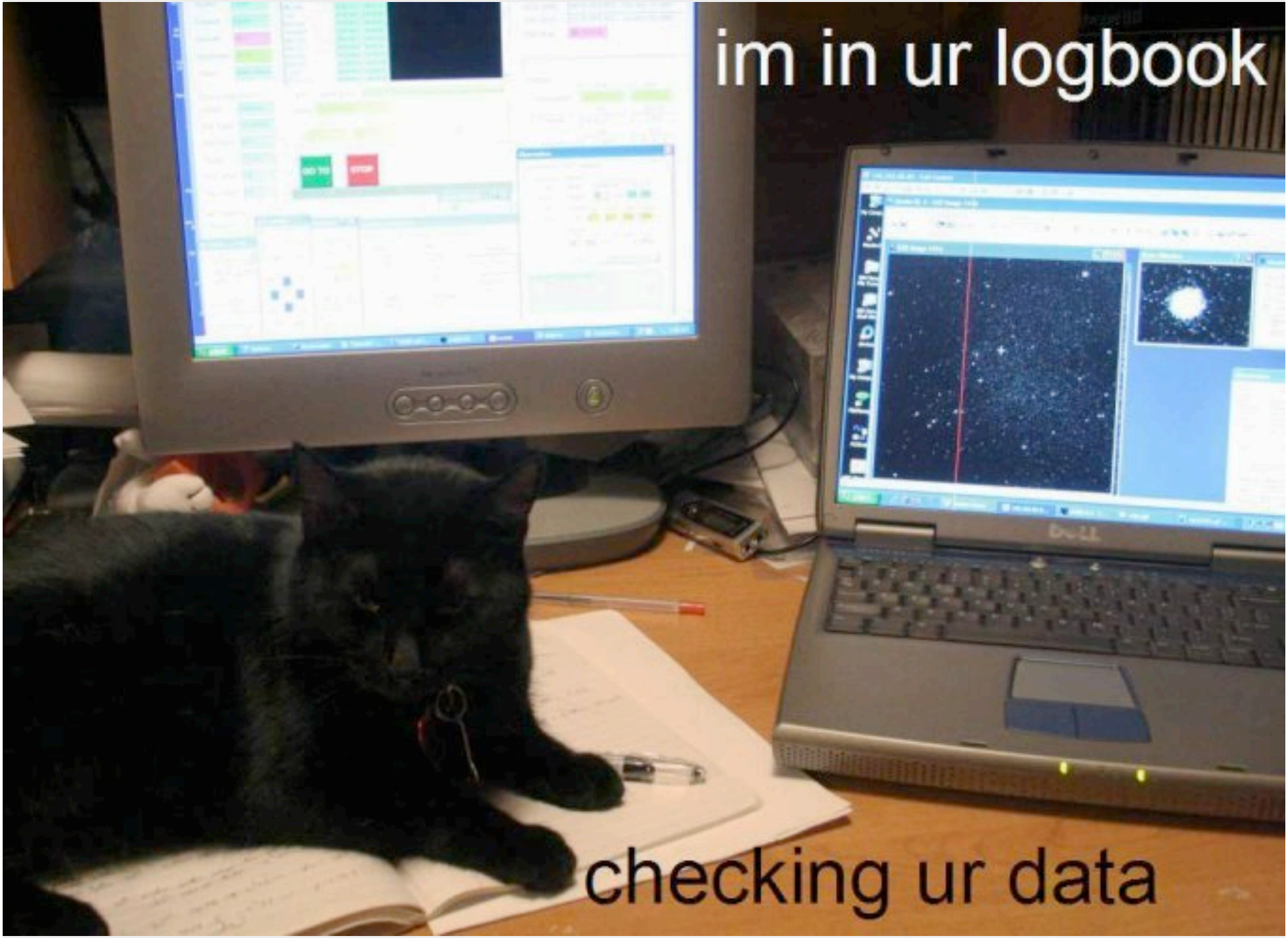
- Exploits the cognitive abilities of **Human Computation**!
- Novel mode of data collection:
  - Citizen Science! = Volunteer Science
  - *e.g.,* VGI = Volunteer Geographic Information (Goodchild '07)
  - *e.g.,* Galaxy Zoo @ http://www.galaxyzoo.org/
- Citizen science refers to the involvement of volunteer non-professionals in the research enterprise.
- The Citizen Science experience …
  - must be engaging,
  - must work with real scientific data/information,
  - must not be busy-work,
  - **must address authentic science research questions** that are beyond the capacity of science teams and enterprises, and
  - must involve the scientists.

Examples of Volunteer Science

- AAVSO (Amer. Assoc. of Variable Star Observers)
- Audubon Bird Counts
- Project Budburst
- Stardust@Home
- VGI (Volunteer Geographic Information)
- CoCoRaHS (Community Collaborative Rain, Hail and Snow network)
- Galaxy Zoo **(~20 refereed pubs so far…)**
- Zooniverse (buffet of Zoos)
- U-Science (semantic science 2.0) [ref: Borne 2009]
  - includes Biodas.org, Wikiproteins, HPKB, AstroDAS
  - **U**biquitous, **U**ser-oriented, **U**ser-led, **U**niversal, **U**ntethered, **You**-centric **Science**

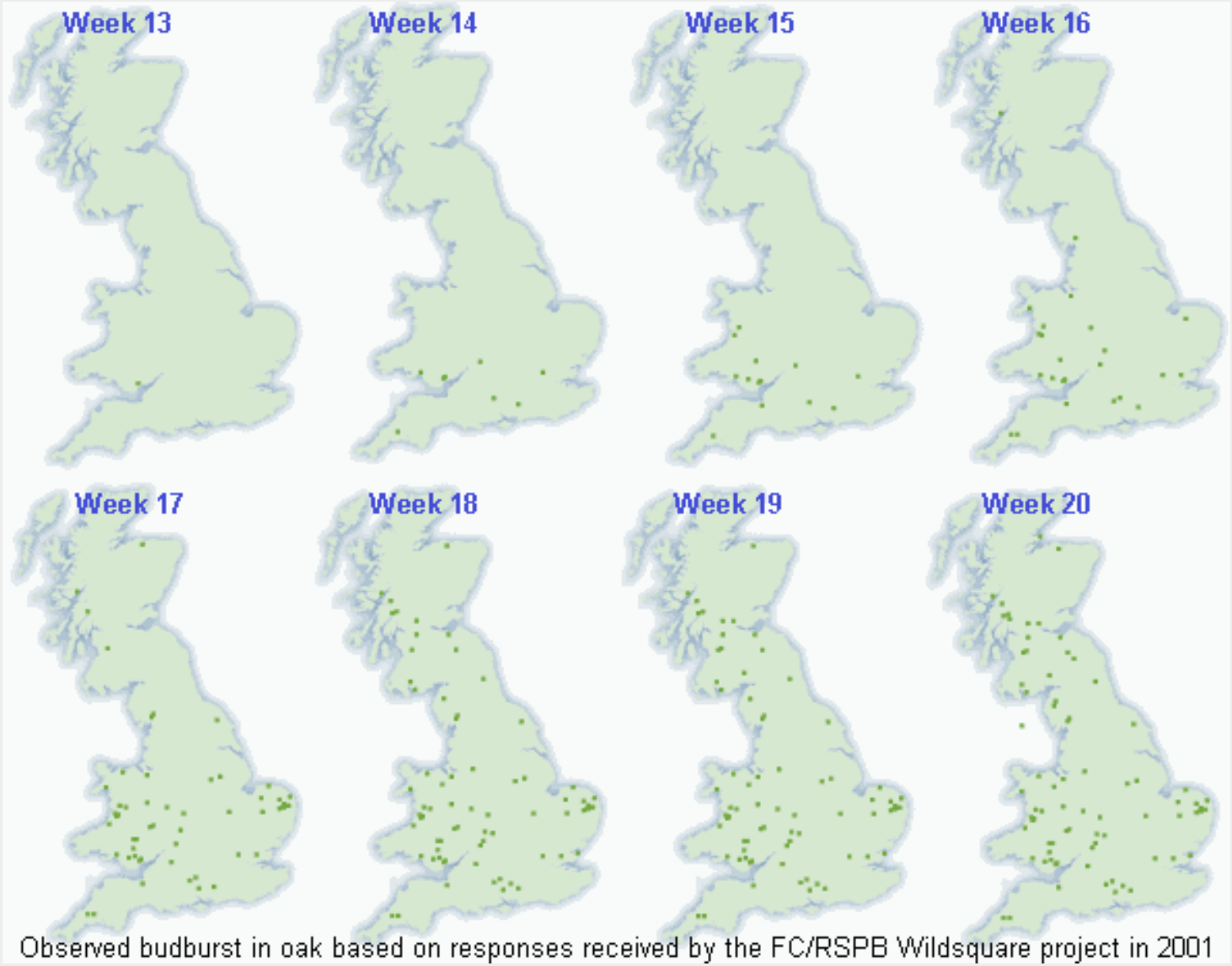
Anybody can participate and contribute to the science...

"On the Internet, nobody knows you're a dog"


im in ur logbook
checking ur data

Project BudBurst: http://www.budburst.org/

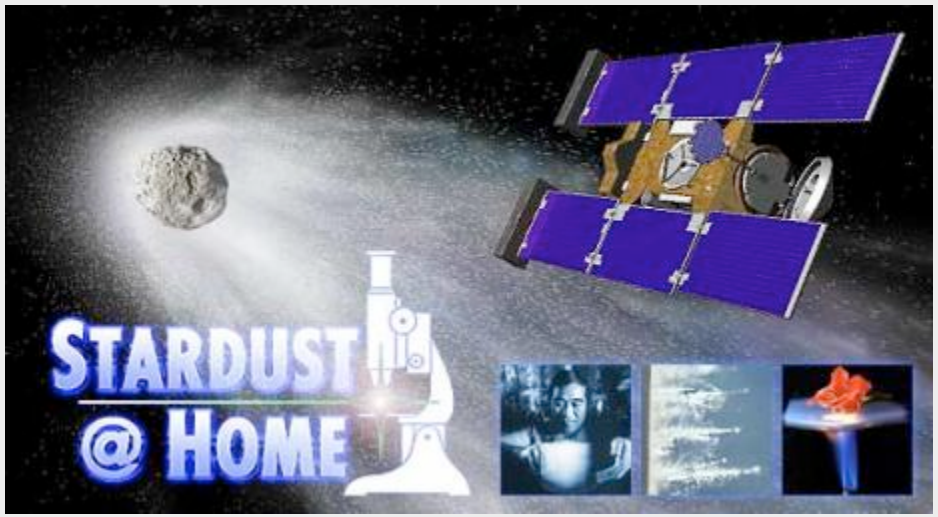
You can help to track climate change in your own backyard by tracking the budding of new springtime flowers.



Observed budburst in oak based on responses received by the FC/RSPB Wildsquare project in 2001

Stardust@home: http://stardustathome.ssl.berkeley.edu/

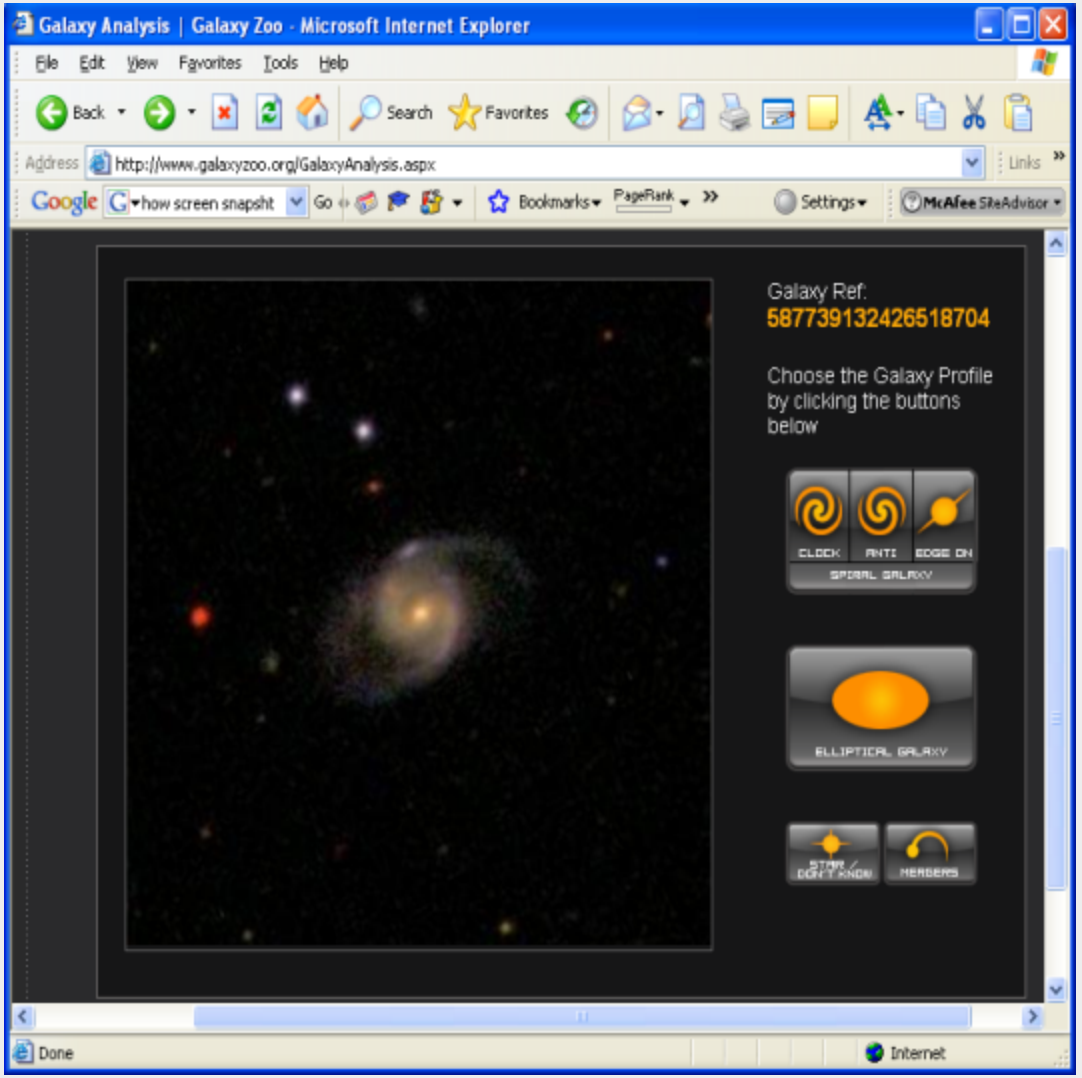
You can help to find star dust particles in aerogel returned from space by NASA's Stardust interplanetary probe.



GalaxyZoo: http://www.galaxyzoo.org/

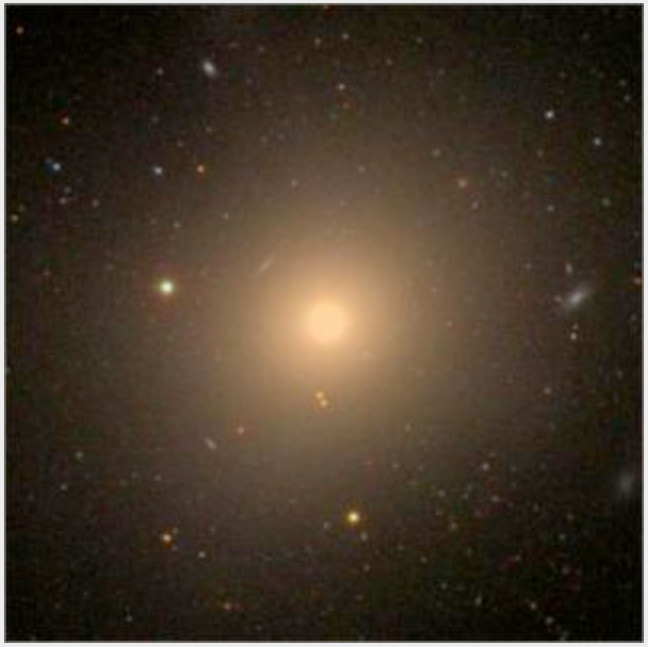
You can help us to classify a million galaxies!

- **"Welcome to GalaxyZoo**, the project which harnesses the power of the internet - and your brain - to classify a million galaxies. By taking part, you'll not only be contributing to scientific research, but you'll view parts of the Universe that literally no-one has ever seen before and get a sense of the glorious diversity of galaxies that pepper the sky."
- **"Why do we need you?** – The simple answer is that the human brain is much better at recognizing patterns than a computer can ever be. Any computer program we write to sort our galaxies into categories would do a reasonable job, but it would also inevitably throw out the unusual, the weird and the wonderful. To rescue these interesting systems which have a story to tell, we need you."



GalaxyZoo helps scientists by engaging the public (hundreds of thousands of us) to classify millions of galaxies:

Is it a Spiral Galaxy or Elliptical Galaxy?

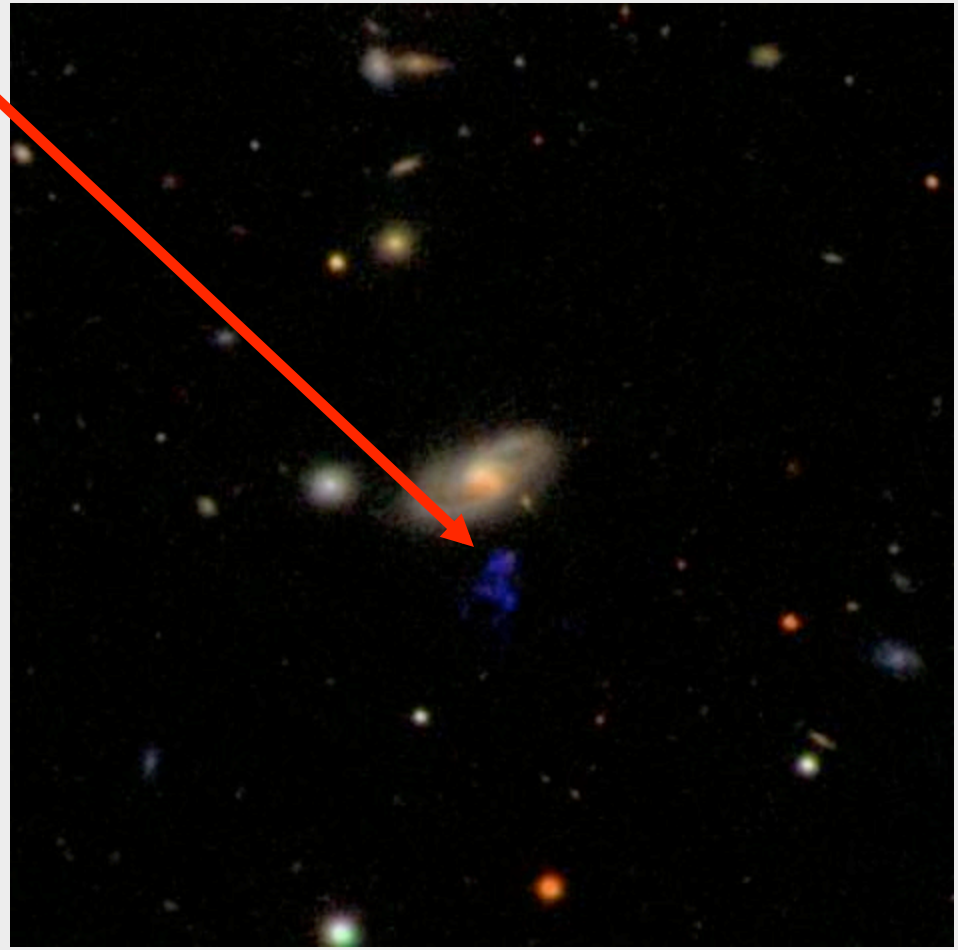


GALAXY ZOO 2
Home
How To Take Part
My Galaxies
Contact Us
Profile
Logout
INVERT GALAXY IMAGE
ADD TO MY FAVOURITES
Classify Galaxies
Answer the question below using the buttons provided.
Is the galaxy simply smooth and rounded, with no sign of a disk?
Smooth
Features or disk
Star or artifact
Need help?

GALAXY ZOO 2
Home
How To Take Part
My Galaxies
Contact Us
Profile
Logout
INVERT GALAXY IMAGE
ADD TO MY FAVOURITES
Classify Galaxies
Answer the question below using the buttons provided.
Is the galaxy simply smooth and rounded, with no sign of a disk?
Smooth
Features or disk
Star or artifact
Need help?

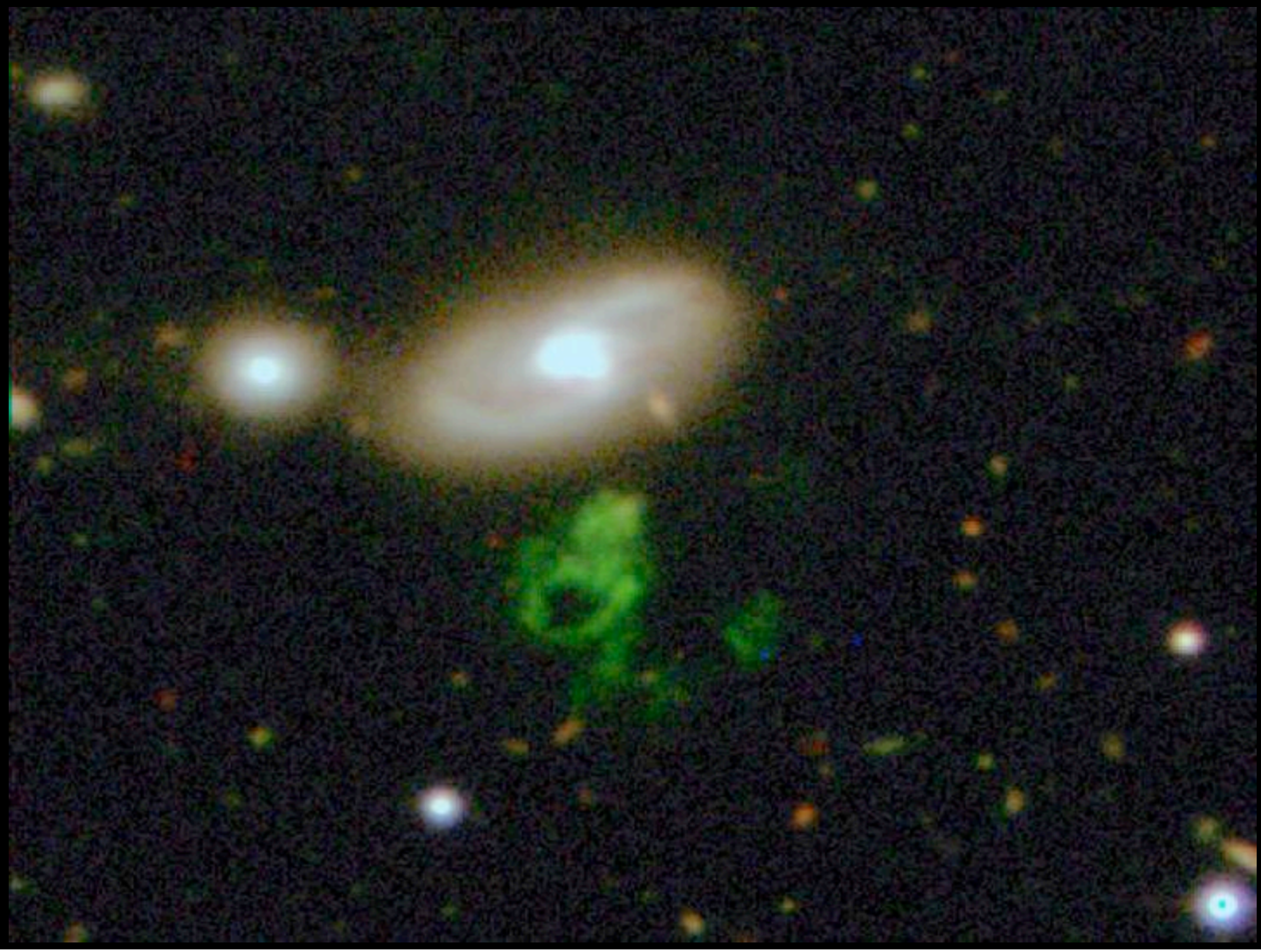
Hanny's Voorwerp: Hanny's Object – found by a school teacher in Holland

- **What is that green thing?** A volunteer sky enthusiast surfing through online Galaxy Zoo images has discovered something really strange. The mystery object is unusually green, not of any clear galaxy type, and situated below relatively normal looking spiral galaxy IC 2497. Dutch schoolteacher Hanny van Arkel, discovered the strange green "voorwerp" (Dutch for "object") last year. The Galaxy Zoo project encourages sky enthusiasts to browse through SDSS images and classify galaxy types. Now known popularly as Hanny's Voorwerp, subsequent observations have shown that the mysterious green blob has the same distance as neighboring galaxy IC 2497. Research is ongoing, but one leading hypothesis holds that Hanny's Voorwerp is a small galaxy that acts like a large reflection nebula, showing the reflected light of a bright quasar event that was visible in the center of IC 2497 about 100,000 years ago. Pictured above, Hanny's Voorwerp was imaged recently by the 2.5-meter Isaac Newton Telescope in the Canary Islands by Dan Smith, Peter Herbert and Chris Lintott (Univ. Hertfordshire). Other collaboration members include Matt Jarvis, Kevin Schawinski, and William Keel.



True color picture of Hanny's Voorwerp:

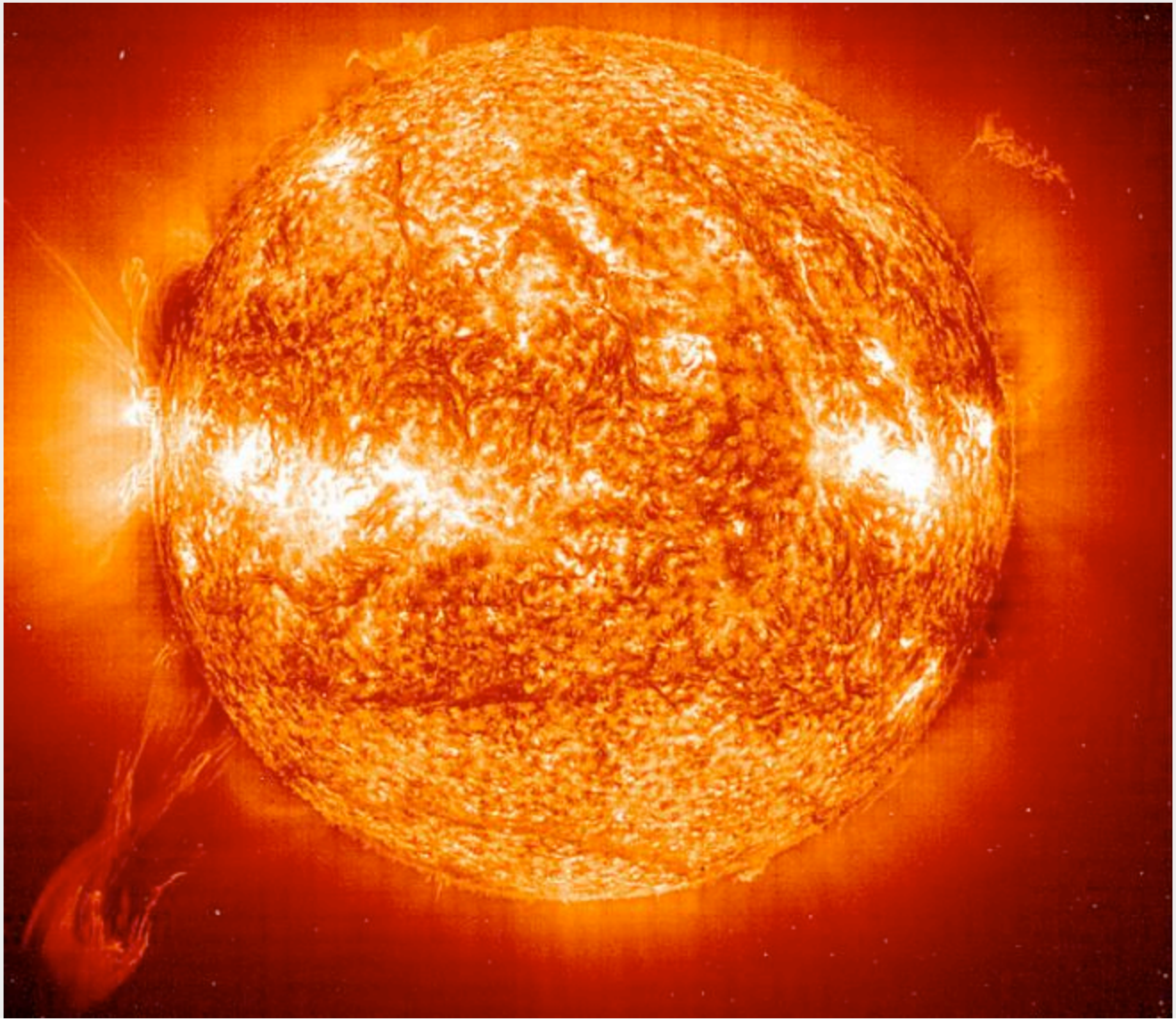
Hanny's Object – the green blob is probably a light echo from an old Quasar that burned out 100,000 years ago



Other big Science Projects can use this idea too ...

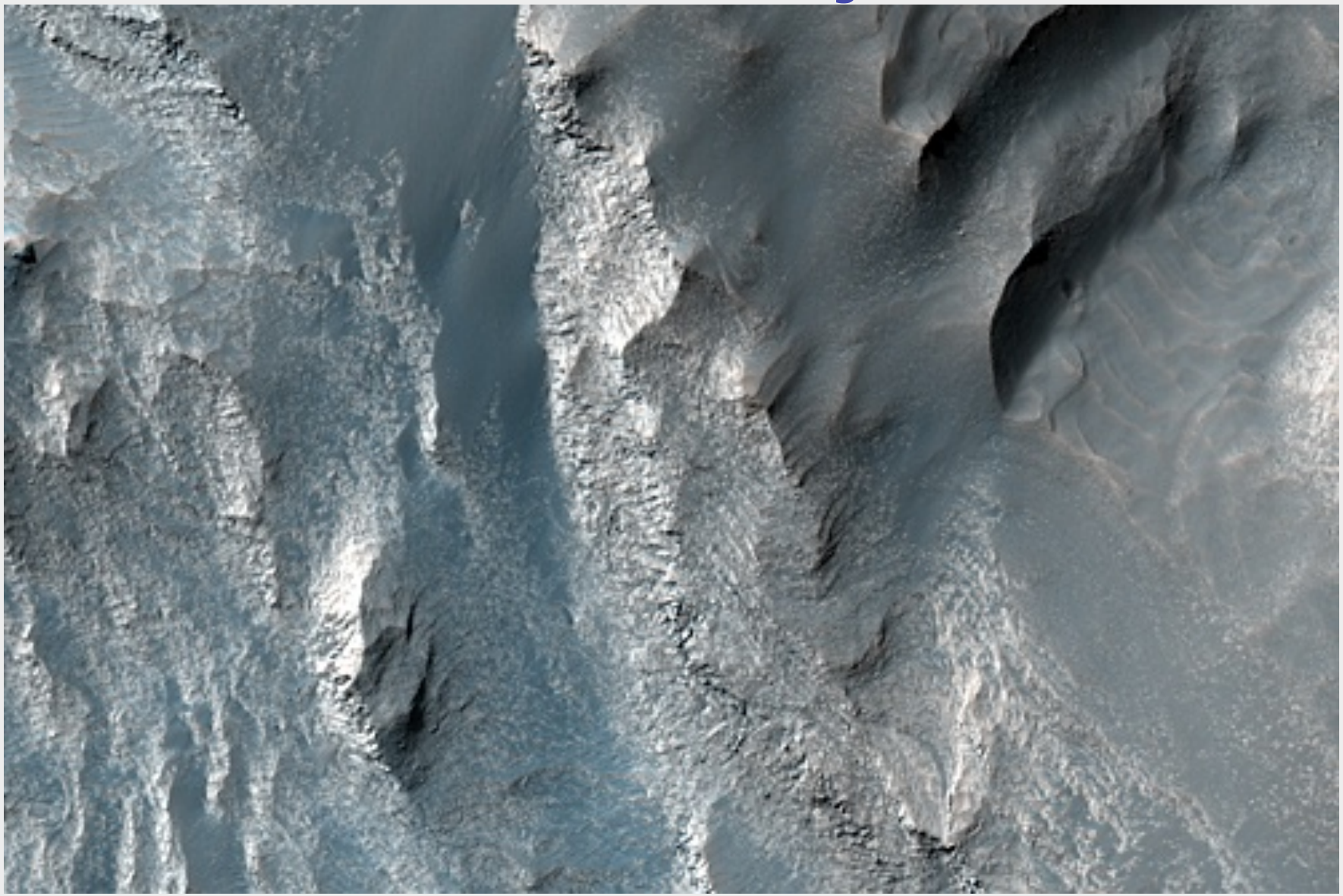
NASA's Solar Dynamics Observatory (SDO):

SDO will generate thousands of images every day, with enormous detail – Citizen Science helps in identifying and sorting out all of the amazing dynamic features on the Sun.



NASA's Mars Reconnaissance Orbiter:

Citizen Science can help to classify and label millions of such images

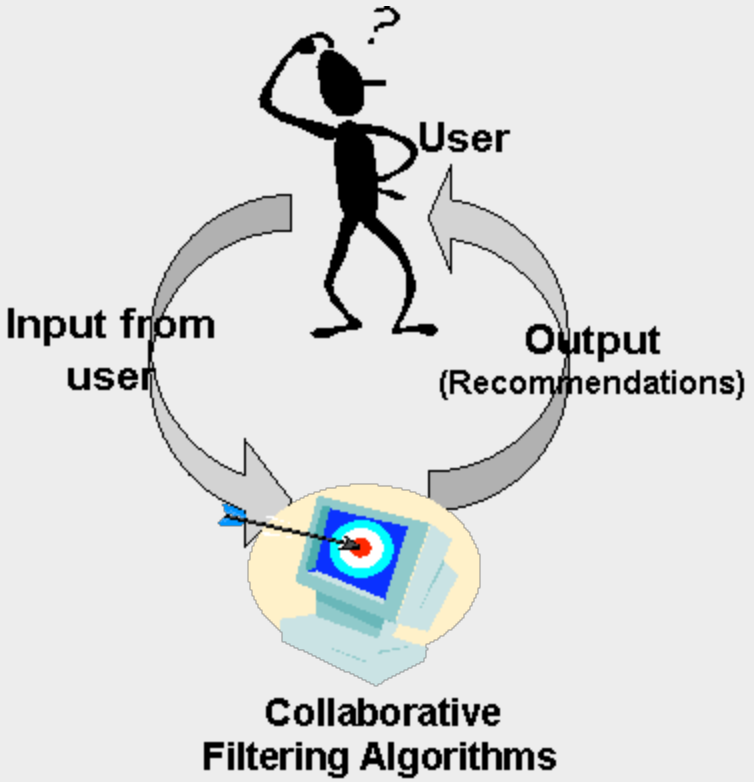


NASA's Lunar Reconnaissance Orbiter:

Citizen Science can help to classify and label all of the different features and regions of the moon

What you say helps others !

- Many types of analysis still require human input to complete.
- Machines are not yet as good as people when it comes to “visual” analysis, anomaly detection, and pattern recognition & classification.
- Large numbers of contributors make light work.
- Eventually, it may be possible to “teach” computers how to perform these tasks following the example of humans: MACHINE LEARNING.
- In U-Science, database end-users help scientific research and discovery by annotating the data with new knowledge!



Classify Galaxies
Report your local weather
Count birds
Count mushrooms
Count bees

Tags produce a new data flood

- **Tagging enables semantic data fusion & integration, and knowledge acquisition / representation / sharing.**
- User-contributed content adds **more** data to the data flood.
- Example – Galaxy Zoo project:
  - ~260,000 participants (*and growing*)
  - ~1 million galaxies have been labeled (classified)
  - ~180 million classifications have been collected
- Tagging is applicable to any data source, including document repositories – adding lightweight semantics to the data repository (taxonomies, folksonomies, annotations)
- Reference: *"TagLearner: A P2P Classifier Learning System from Collaboratively Tagged Text Documents",* Dutta, Zhu, Mahule, Kargupta, Borne, Lauth, Holz, & Heyer, 2009 ICDM paper.

Outline

- Astronomy example: the LSST Project
  - Classification of millions of real-time events
  - Classification of billions of galaxies
- Human Computation
- U-Science
- The Zooniverse Project

The Zooniverse* :
Advancing Science through User-Guided Learning in Massive Data Streams

* NSF CDI funded program @ http://zooniverse.org

The Zooniverse

http://zooniverse.org/

- New funded NSF CDI grant (PI: L.Fortson, Adler Planetarium; co-PI J. Wallin & collaborator K.Borne, GMU; & collaborators at Oxford U)
- Building a framework for new Citizen Science projects, including user-based research tools
- Science domains:
  - Astronomy (Galaxy Merger Zoo)
  - The Moon (Lunar Reconnaissance Orbiter)
  - The Sun (STEREO dual spacecraft)
  - Egyptology (the Papyri Project)
  - and more (… accepting proposals from community)

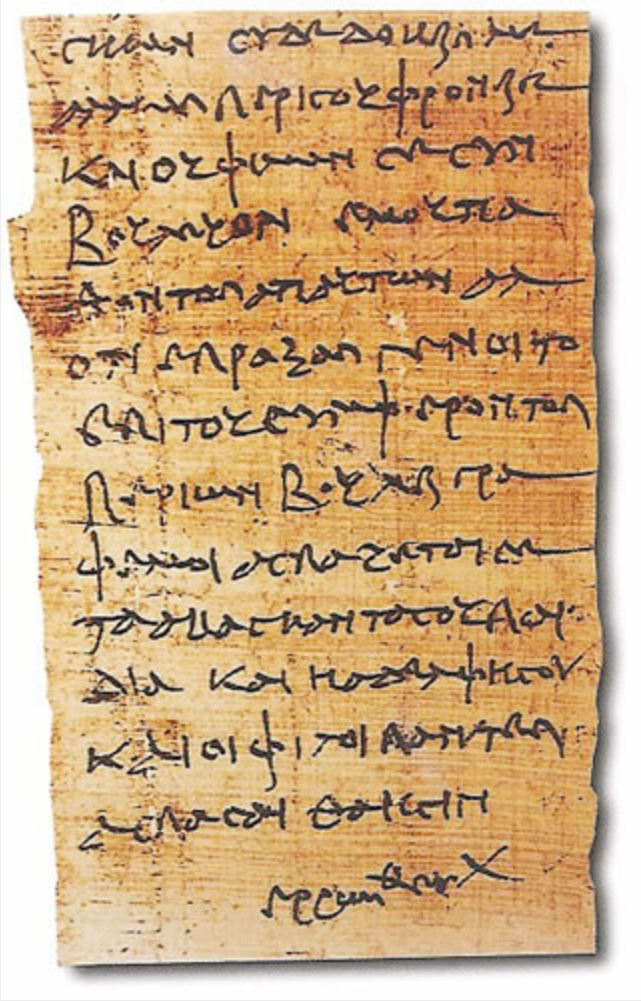
The Zooniverse: a Buffet of Zoos

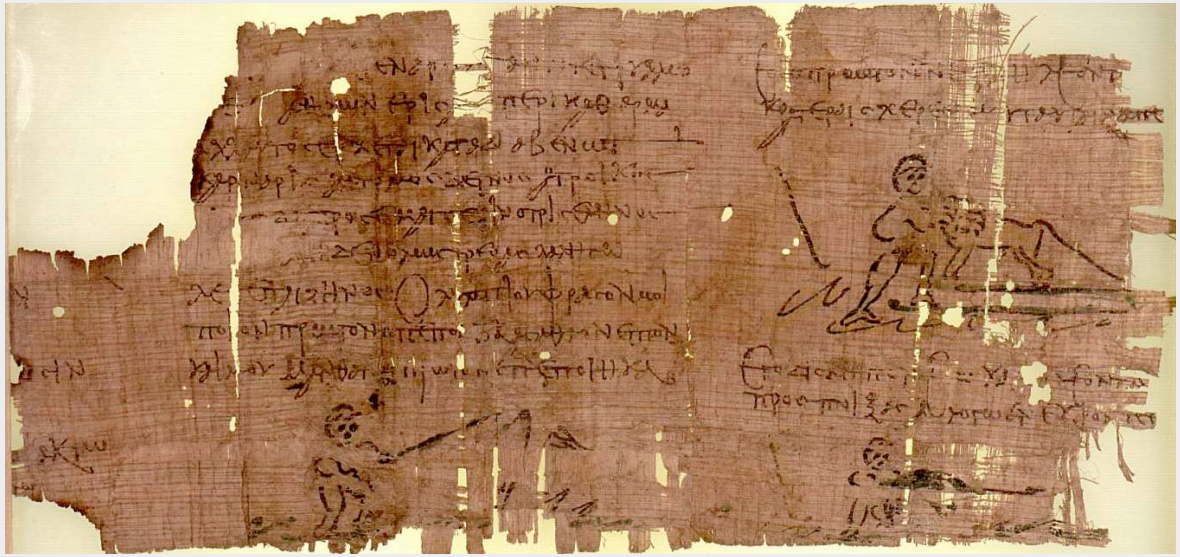
http://zooniverse.org/

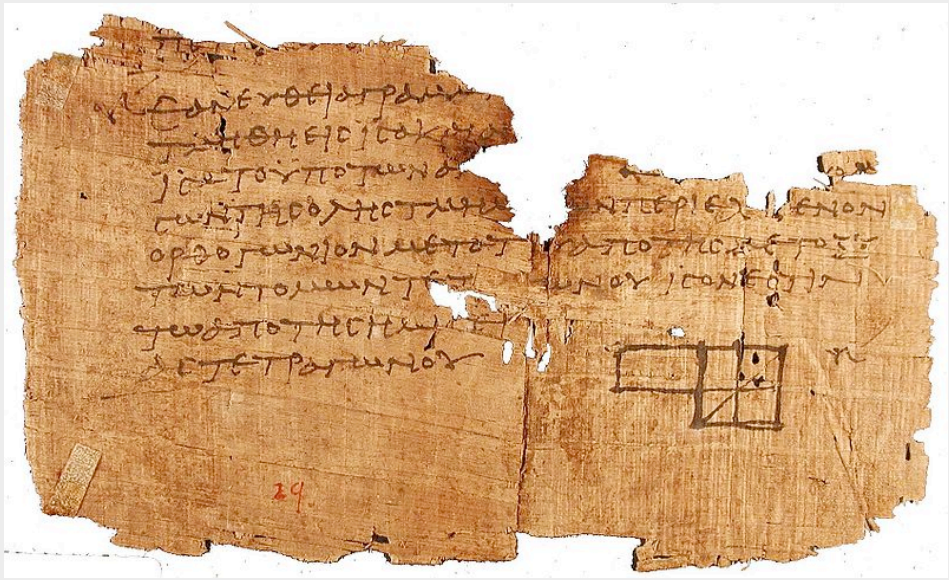
- Galaxy Zoo project (released July 2007):
  - http://www.galaxyzoo.org/
  - Classify galaxies (Spiral, Elliptical, Merger, or image artifact)
- Galaxy Merger Zoo (release November 2009)
  - http://mergers.galaxyzoo.org/
  - Run numerical simulations to find best model to match a real merger
  - One new merger every day
- The Hunt for Supernovae (released December 2009)
  - http://supernova.galaxyzoo.org/
  - Real-time event detection and classification
- Solar Storm Watch (released March 2010)
  - http://solarstormwatch.com/
  - Spot solar storms (CMEs) in near real-time

Egyptology (the Papyri Project)

Oxyrhynchus Papyri Project @ http://www.papyrology.ox.ac.uk/







Key Feature of Zooniverse:
Data mining from the volunteer-contributed labels

- Train the automated pipeline classifiers with:
  - Improved classification algorithms
  - Better identification of anomalies
  - Fewer classification errors
- Millions of training examples
- Hundreds of millions of class labels
- Statistics deluxe! …
  - Users (see paper: http://arxiv.org/abs/0909.2925)
  - Uncertainty quantification
  - Classification certainty vs. Classification dispersion

Challenge Problems

- **Zooniverse Data Mining (Machine Learning) Challenge Problems (2011-2013)**

Other similar examples:

- KDD cups
- Netflix Prize (#1 and #2)
- GREAT08 Challenge
- Digging into Data Challenge 2009 (diggingintodata.org)
- Transportation challenge problems
- KD2u.org – knowledge discovery from challenge data sets

What's next?

We have described some remarkable results from the field of citizen science, based upon work with static databases.

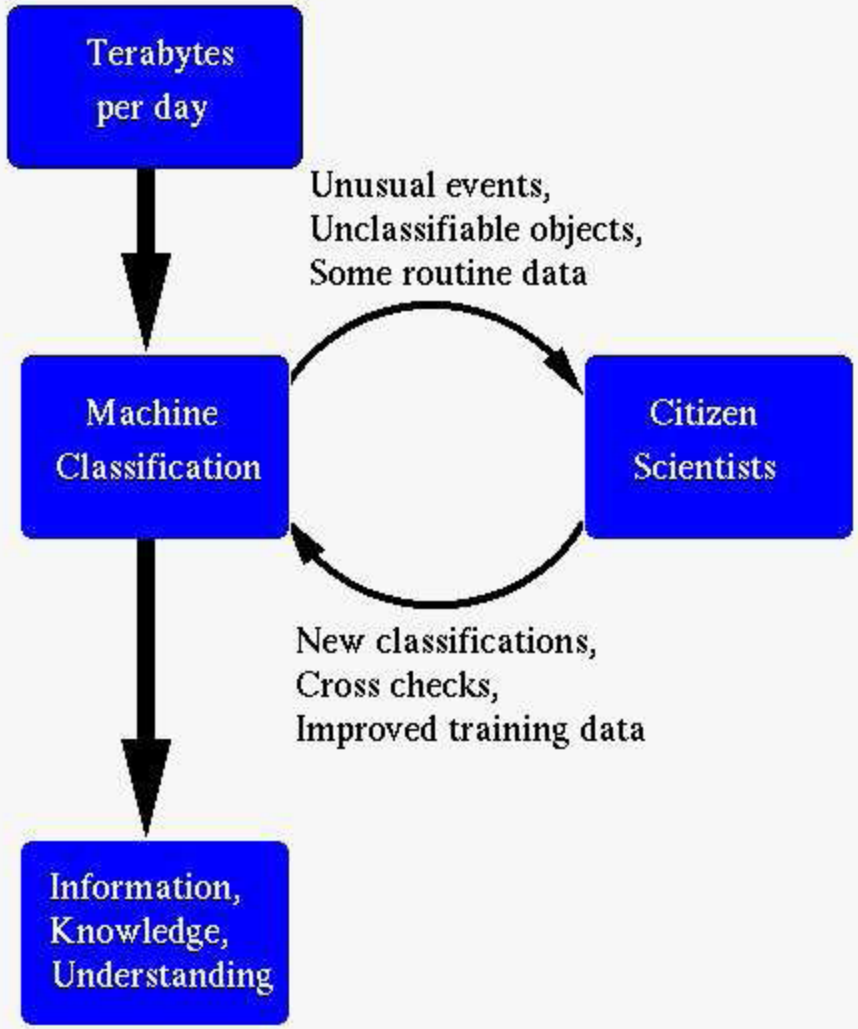
We envision the eventual application of this emerging computational resource to the problem of massive stream mining for scientific discovery.

The High Energy Physics Paradigm, compared with the Large Synoptic Survey Telescope (LSST)

- LHC (Large Hadron Collider) will generate particle tracks at the rate of ~1 Petabyte/sec.
- To cope with this enormous data rate, the data stream is mined in real-time (single-pass), with only about 1 KB of the data being preserved (= the "interesting" tracks).
- LSST will generate a slower rate data stream of "only" 30 TB per night, but it is still cognitively challenging.
  - Data has *inertia* (due to I/O and communication bottlenecks), so mine the data as it moves through the cloud.
  - As the data move from camera/detector to the archive (data management) system, present the data (images and other modalities) in streaming mode to a waiting audience of interested volunteers – ready to tag, annotate, and discover!

Challenge Areas and The Future Man-Machine Partnership

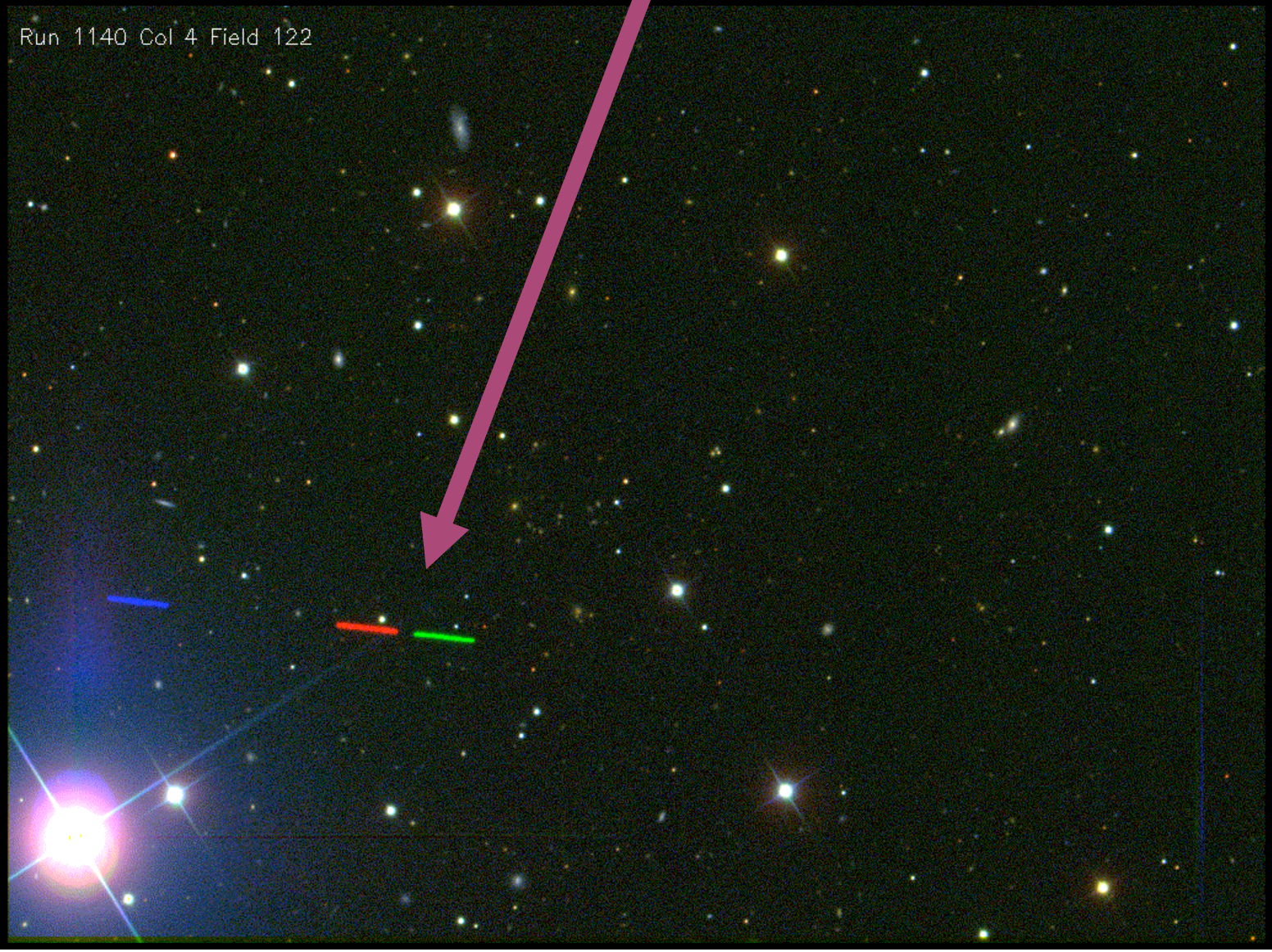
- Data volumes
- Scalability
- Real-time analytics
- One-pass data stream
- Trust



Conclusion: approaches to addressing the science data flood

- ***X-Informatics*** (*e.g.,* X = Bio, Geo, Astro, …):
  - addresses the scientific data lifecycle challenges in the era of data-intensive science and the data flood
  - defines lightweight ontologies, semantics, taxonomies, concepts, content descriptors for a science domain
  - for the purpose of organizing, accessing, searching, fusing, integrating, mining, and analyzing massive data repositories.
- User-guided informatics-powered ***Analytics in the Cloud***:
  - Human computation (*e.g.,* tagging, labeling, classification)
    - characterized by enormous cognitive capacity and pattern recognition efficiency
  - U-Science (Semantic e-Science **and** Volunteer Citizen Science)
  - Tagging everything, everywhere

Maybe you can help us to find the next Killer Asteroid!



Related References

- Borne (2009): *"**U-Science**"*, http://essi.gsfc.nasa.gov/pdf/Borne2.pdf
- Borne, Jacoby, …, Wallin (2009): *"**The Revolution in Astronomy Education: Data Science for the Masses**"*, http://arxiv.org/abs/0909.3895
- Borne (2009): *"**Astroinformatics: A 21st Century Approach to Astronomy**"*, http://arxiv.org/abs/0909.3892
- Dutta, Zhu, Mahule, Kargupta, Borne, Lauth, Holz, & Heyer (2009): *"**TagLearner: A P2P Classifier Learning System from Collaboratively Tagged Text Documents**"*, accepted paper for ICDM-2009.
- M. F. Goodchild (2007): *"**Citizens as Sensors: the World of Volunteered Geography**"*, GeoJournal, 69, pp. 211-221.
- Lintott et al. (2009): *"**Galaxy Zoo: Hanny's Voorwerp', a quasar light echo?**"*, http://arxiv.org/abs/0906.5304
- Raddick et al. (2009): *"**Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers**"*, http://arxiv.org/abs/0909.2925
- Raddick, Bracey, Carney, Gyuk, Borne, Wallin, & Jacoby (2009): *"**Citizen Science: Status and Research Directions for the Coming Decade**"*, http://www8.nationalacademies.org/astro2010/DetailFileDisplay.aspx?id=454

Related References

- Borne (2009): *"**U-Science**",* http://essi.gsfc.nasa.gov/pdf/Borne2.pdf
- Borne, Jacoby, …, Wallin (2009): *"**The Revolution in Astronomy Education: Data Science for the Masses**",* http://arxiv.org/abs/0909.3895
- Borne (2009): *"**Astroinformatics: A 21st Century Approach to Astronomy**",* http://arxiv.org/abs/0909.3892
- Dutta, Zhu, Mahule, Kargupta, Borne, Lauth, Holz, & Heyer (2009): *"**TagLearner: A P2P Classifier Learning System from Collaboratively Tagged Text Documents**",* accepted paper for ICDM-2009.
- M. F. Goodchild (2007): *"**Citizens as Sensors: the World of Volunteered Geography**",* GeoJournal, 69, pp. 211-221.
- Lintott et al. (2009): *"**Galaxy Zoo: 'Hanny's Voorwerp', a quasar light echo?**",* http://arxiv.org/abs/0906.5304
- Raddick et al. (2009): *"**Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers**",* http://arxiv.org/abs/0909.2925
- Raddick, Bracey, Carney, Gyuk, Borne, Wallin, & Jacoby (2009): *"**Citizen Science: Status and Research Directions for the Coming Decade**",* http://www8.nationalacademies.org/astro2010/DetailFileDisplay.aspx?id=454