Author(s): August E. Evrard, PhD. 2010

Citation Key

for more information see: http://open.umich.edu/wiki/CitationPolicy

Use + Share + Adapt

{ Content the copyright holder, author, or law permits you to use, share and adapt. }

PD-GOV **Public Domain – Government**: Works that are produced by the U.S. Government. (17 USC § 105)

PD-EXP **Public Domain – Expired**: Works that are no longer protected due to an expired copyright term.

PD-SELF **Public Domain – Self Dedicated**: Works that a copyright holder has dedicated to the public domain.

(cc) ZERO **Creative Commons – Zero Waiver**

(cc) BY **Creative Commons – Attribution License**

(cc) BY-SA **Creative Commons – Attribution Share Alike License**

(cc) BY-NC **Creative Commons – Attribution Noncommercial License**

(cc) BY-NC-SA **Creative Commons – Attribution Noncommercial Share Alike License**

GNU-FDL **GNU – Free Documentation License**

Make Your Own Assessment

{ Content Open.Michigan believes can be used, shared, and adapted because it is ineligible for copyright. }

PD-INEL **Public Domain – Ineligible**: Works that are ineligible for copyright protection in the U.S. (17 USC § 102(b)) *laws in your jurisdiction may differ

{ Content Open.Michigan has used under a Fair Use determination. }

FAIR USE **Fair Use**: Use of works that is determined to be Fair consistent with the U.S. Copyright Act. (17 USC § 107) *laws in your jurisdiction may differ

Our determination **DOES NOT** mean that all uses of this 3rd-party content are Fair Uses and we **DO NOT** guarantee that your use of the content is Fair.

To use this content you should **do your own independent analysis** to determine whether or not your use will be Fair.

Cyberscience: Computational Science and the Rise of the Fourth Paradigm

Computational Science and the Rise of the Fourth Paradigm

Honors 352, Class#0.3

August E. (Gus) Evrard, PhD

Fall 2010

University of Michigan

today

* reading quiz

* lecture: broad themes from 4th P, Sec 1,2 reading
  + Cosma as a specific example

* discussion: forming groups and choosing group projects

* reminder: first (wordpress) blog entry due next **Monday 9am**

  choose your theme:

  What has scientific computing done for society?

  What interests me most about scientific computing is ...

science is evolving

- ***basic (natural) sciences*** drove the first three paradigms
  - physics, chemistry, biology, astronomy, etc.
  - many important, fundamental questions:
    nature of matter, origin of life, evolution, …

  particle physics = purist (?) form:
  standard model of particle physics (theory)
  + collider accelerators (experiment)
  + lattice quantum chromodynamics (simulation)

- ***solving societal problems*** under 4th P requires broad scope: *natural + social sciences + engineering*

  healthcare
  climate
  energy
  agriculture
  social justice
  …

> *With science moving toward being computational and data based, key technology challenges include the need to better capture, analyze, model, and visualize scientific information. The ultimate goal is to aid scientists, researchers, policymakers, and the general public in making informed decisions.*

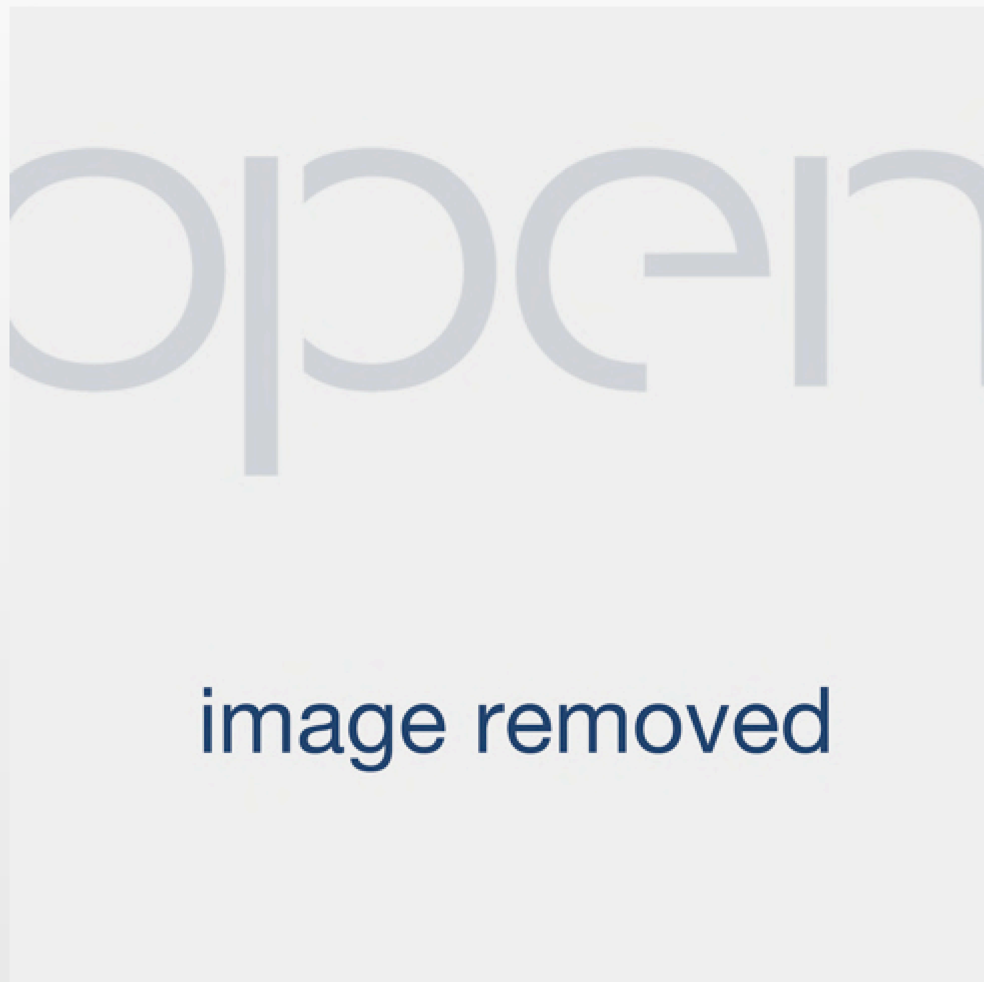
Dan Fay, *4th Paradigm*, p. 3

science is evolving

"We seek solutions. We don't seek – dare I say this? – just scientific papers anymore."

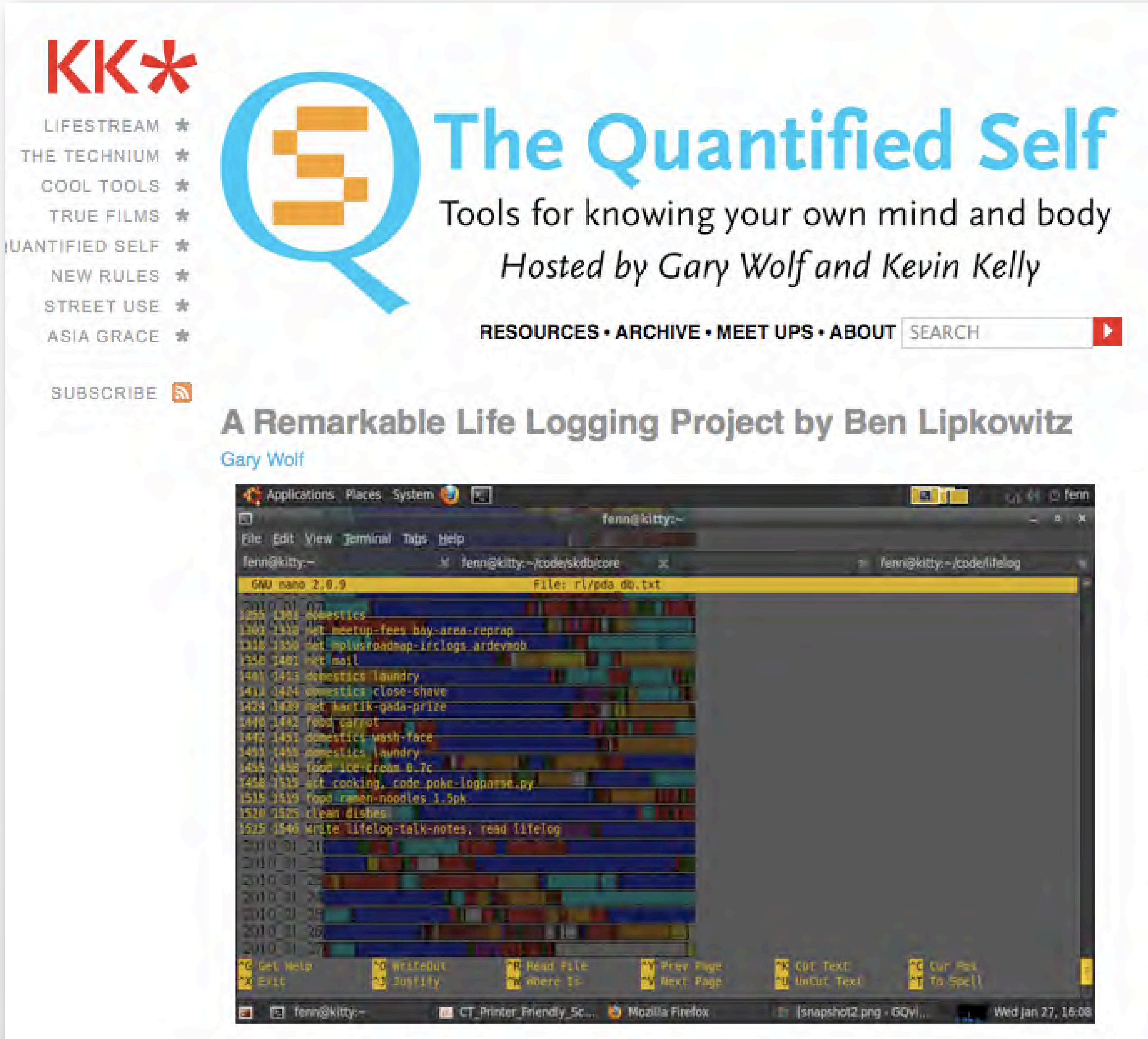
 R. DelVecchio, "UC Berkeley: Panel looks at control of emissions" San Francisco Chronicle, 27 March 2007

opportunities for data capture abound



Please see original image of New York Times article at
http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html

data capture for *self-quantification*

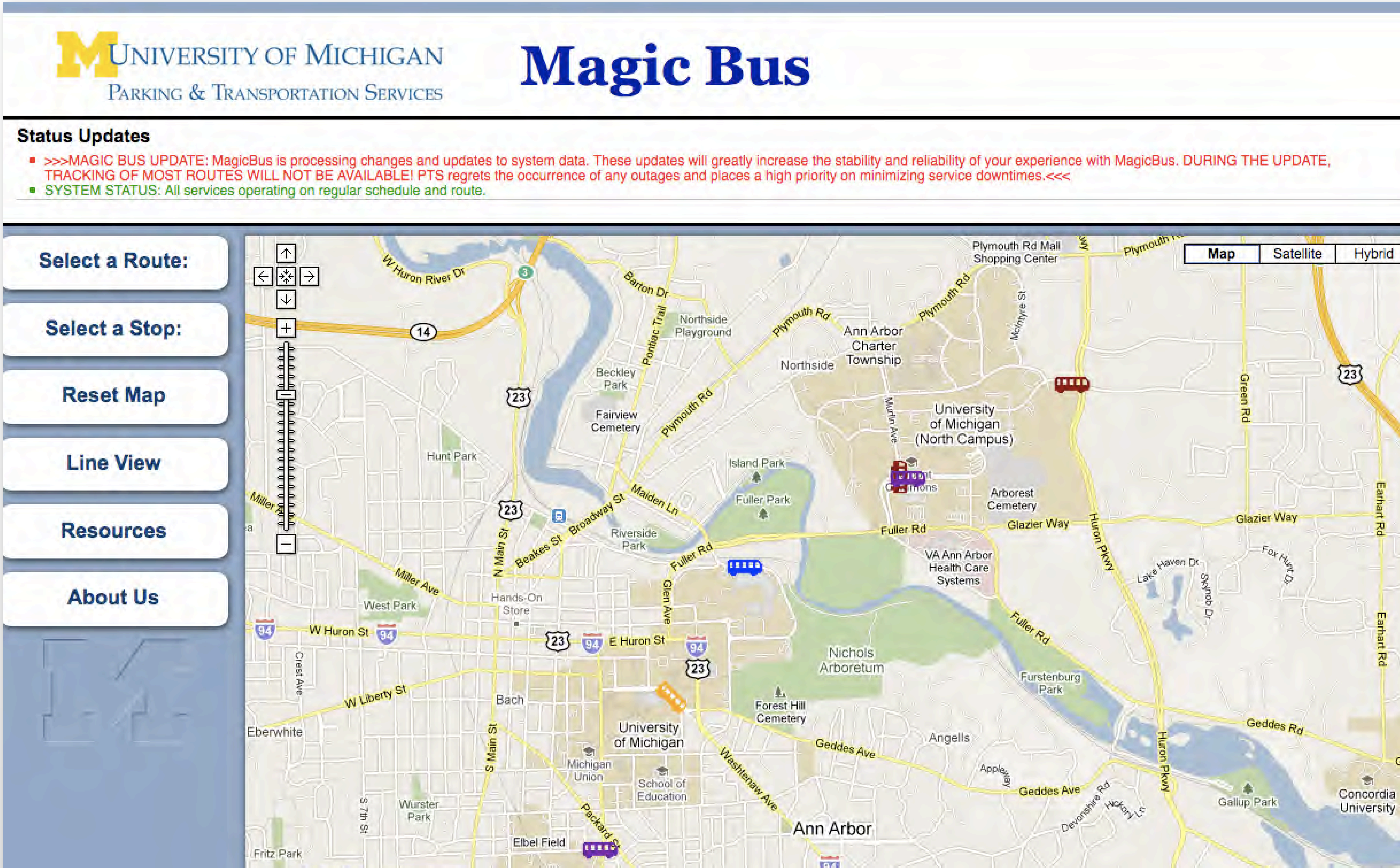


Ben started self-tracking when he wondered how much time he actually spent doing his roommate's dishes. He estimated that he spent an hour a day. (The true amount, he learned, was about 20 minutes.) This led him to track more things. This led him to track everything. ***Ben considers this project a scientific investigation,*** but is a type of science inspired by Buckminster Fuller, aka "Guinea Pig B," rather than by a more conventional academic tradition.

It is easy to imagine somebody watching this video and saying: "well, that's a full scale geek-out that has nothing to do with me." But I think this would be an error. Ben uses data to find out details about himself, such as how he sleeps. He also uses data to figure out problems in his social life, such as whether the distribution of housework is reasonable. While its true that Ben's method of data gathering is unusable by most people, ***the type of questions he's asking and the conviction that these questions are answerable are a preview of the coming age of self-quantification***.

 http://www.kk.org/quantifiedself/2010/02/a-remarkable-life-logging-proj.php

data capture/broadcast for a better life



 http://mbus.pts.umich.edu/

some broad themes

- data exist from many sources
  - government agencies, research societies
  - large project collaborations (= “big science”)
  - small research groups (= “individual investigators”)
  - amateur scientists
  - general public

- scientists conducting a study typically
  - identify appropriate data sources
  - collect and validate data sets
  - process and generate new data

- current data suffer from
  - lack of data definition conventions within research communities
  - weak incentives: ***“knowledge” > “data”***
  - \`raw’ data is often unpublished (invisible to WWW)
  - what’s worth keeping?

flavors of data-intensive science

- *ab initio* simulations
  - solve governing equations of a dynamical system
- model testing
  - compare theoretical/computational model expectations to observations/experiment
  - perform statistical (likelihood) analysis to confirm/refute model or constrain model parameter values
- empirical studies of natural or social phenomena (*data mining*)
  - correlations among internal properties of a population or system
  - influence of external factors
  - spatial or temporal dependencies
- signal processing
  - separate data into a useful portion – the *signal* – and unwanted noise
  - estimate errors on measured signals

<u>Mature scientific areas involve all of these!</u>

Cosma NSF proposal: a cyberscience *instantiation*

- Dark Energy Survey = `medium-sized' (~40M$) astronomical survey of galaxies
  - follows Sloan Digital Sky Survey (SDSS), ~2000
  - precursor to Large Synoptic Survey Telescope (LSST), ~2020
  - makes happy both Astrophysicists and Cosmologists!
    (A: How did galaxies form? C: How is space expanding?)
  - first survey to use four dark energy methodologies on same data set (CL, WL, BAO, SN)

- collaboration of ~300 observers, theorists, computational scientists
  - sparse funding model: camera and database fully funded, nothing (yet) for analysis
  - proposal seeks to fund small team focused on ***simulating model expectations***

- preparing for survey data analysis requires planning and preparation
  - simulations of cosmic structure formation `dressed' with galaxies => `truth' sky catalog
  - pretend to observe `truth' by adding real-world effects => `observed' sky catalog
    distortions in galaxy images from intervening matter (gravitational lensing)
    distortions from telescope optics and camera electronics
  - process `observed' catalog in same manner as real data
  - compare to `truth' to estimate fidelity of survey processing

- iterative process
  - science teams employ catalogs with different levels of detail
  - multiple simulations to provide modest coverage of model space

Additional Source Information

for more information see: http://open.umich.edu/wiki/CitationPolicy

Slide 3: Scott Beale, "Columbia Supercomputer," http://www.flickr.com/photos/27403767@N00/465519540/, CC: BY-NC-ND 2.0, http://creativecommons.org/licenses/by-nc-nd/2.0/deed.en

Slide 5, Image (top): tsevis, "Barack Obama: A mosaic of people," Flickr, http://www.flickr.com/photos/tsevis/2279253649, CC: BY-NC-SA 2.0, http://creativecommons.org/licenses/by-nc-sa/2.0/deed.en
Slide 5, Quote (bottom): Dan Fay, *4th Paradigm*, p. 3

Slide 6, Image (top): Lawrence Berkeley National Laboratory, "Energy Secretary Chu Visits Berkeley Lab," Flickr, http://www.flickr.com/photos/29648341@N08/4050433770/, CC: BY-NC-ND 2.0, http://creativecommons.org/licenses/by-nc-nd/2.0/deed.en
Slide 6, Quote (bottom): *R. DelVecchio, "UC Berkeley: Panel looks at control of emissions" San Francisco Chronicle, 27 March 2007*

Slide 7: Please see original image of New York Times article at http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html

Slide 8: Please see original image of screenshot and quote from regarding quantification at http://www.kk.org/quantifiedself/2010/02/a-remarkable-life-logging-proj.php

Slide 9: Please see original image of real-time University of Michigan bus tracking program at http://mbus.pts.umich.edu/.

Slide 10: woodleywonderworks, "The Night Lights of the United States (as seen from space)," Flickr, http://www.flickr.com/photos/wwworks/2712986388/, CC: BY 2.0, http://creativecommons.org/licenses/by/2.0/deed.en