High Energy Physics and the Citizen Scientist

Harrison B. Prosper
Florida State University

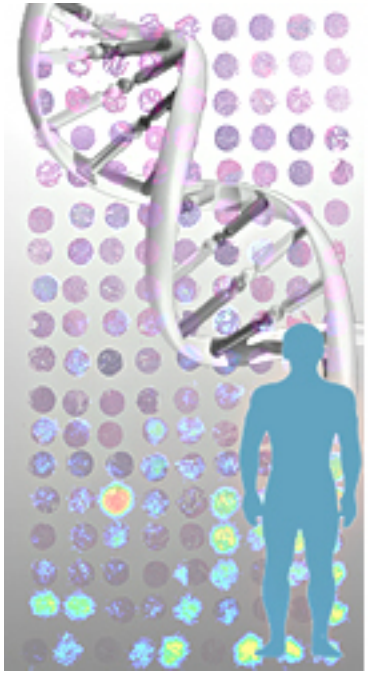
IPNL, Lyon

18 April, 2011

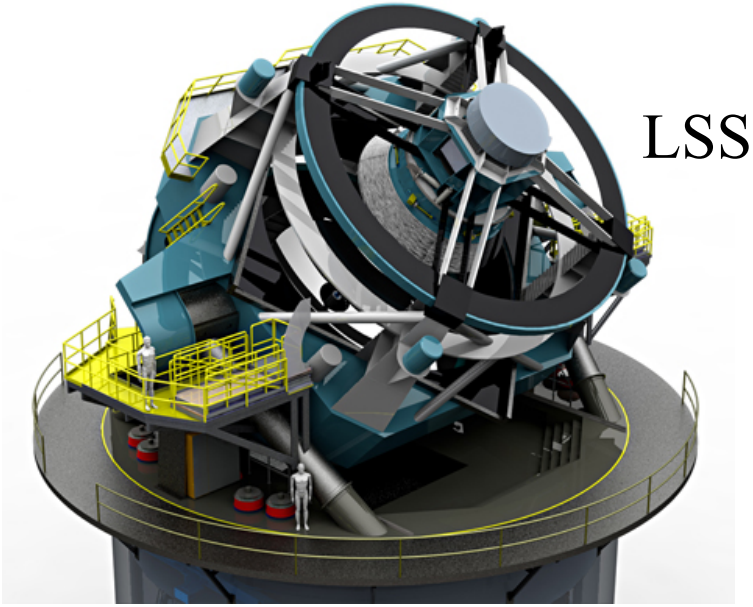
Hanford, Washington

Livingston, Louisiana

LIGO
AdvLIGO
Virgo



US
National
Human
Genome
Project



LSST

Open Data

These are examples of large-scale science projects that have embraced an ***Open Data*** policy.

Indeed, funding of these projects is contingent on their adherence to the policy of

systematic, periodic, data release and public access

See, for example, this report by the Welcome Trust:

THE WELLCOME TRUST

Summary of the Report of the Second International Strategy Meeting on Human Genome Sequencing

Bermuda, 27th February - 2nd March 1997
Sponsored by the Wellcome Trust

Why Open Data?

1. Our future depends on it!

2. More brainpower means more creative science.

   In the age group 15-64 years, there are **4.5 billion brains**. Compare this with the **2500** permanent brains at CERN + its **8000** visiting brains.

3. Data acquired today (at great cost) can yield novel science tomorrow.

4. Data paid by the public are owned by the public.

Why Open HEP Data?

"Reports about data from LEP experiments being increasingly difficult to analyze or even becoming inaccessible are starting to surface."

ICFA Study Group for Data Preservation

DPHEP-2009-001
November 30, 2009

http://www.toonpool.com/cartoons/data%20loss_5574

All Data to All People!

Rocky Kolb (0708.1199v1) identified the following excuses for closed data:

Excuse 1: HEP experiments are too complicated.

Excuse 2: Wrong papers would be published.

Excuse 3: It would be too expensive.

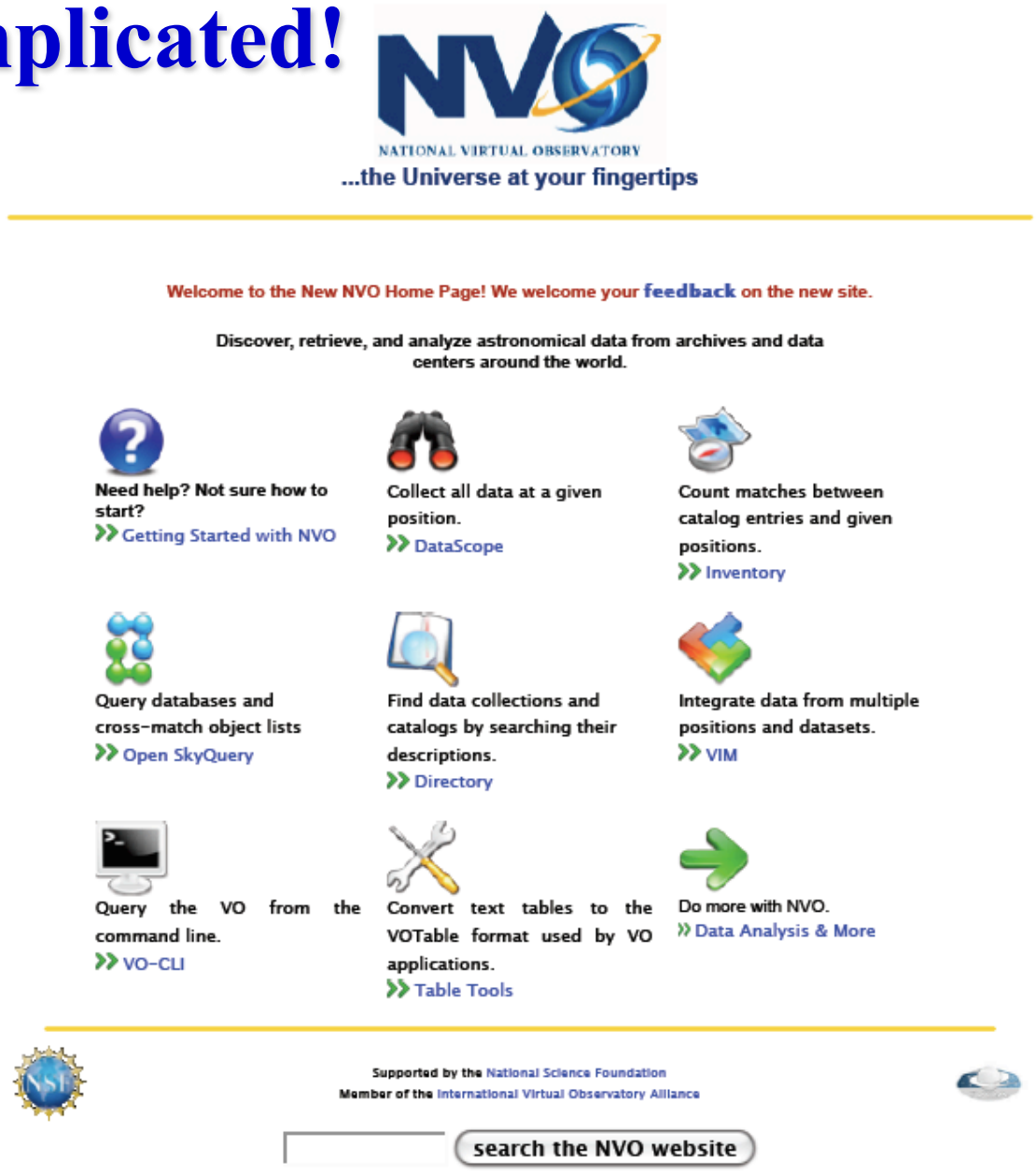
Excuse 4: It wouldn't be fair to the hard-working physicist.

Excuse 1 – Complicated!

The National Virtual Observatory provides open access to data from a wide variety of sources.

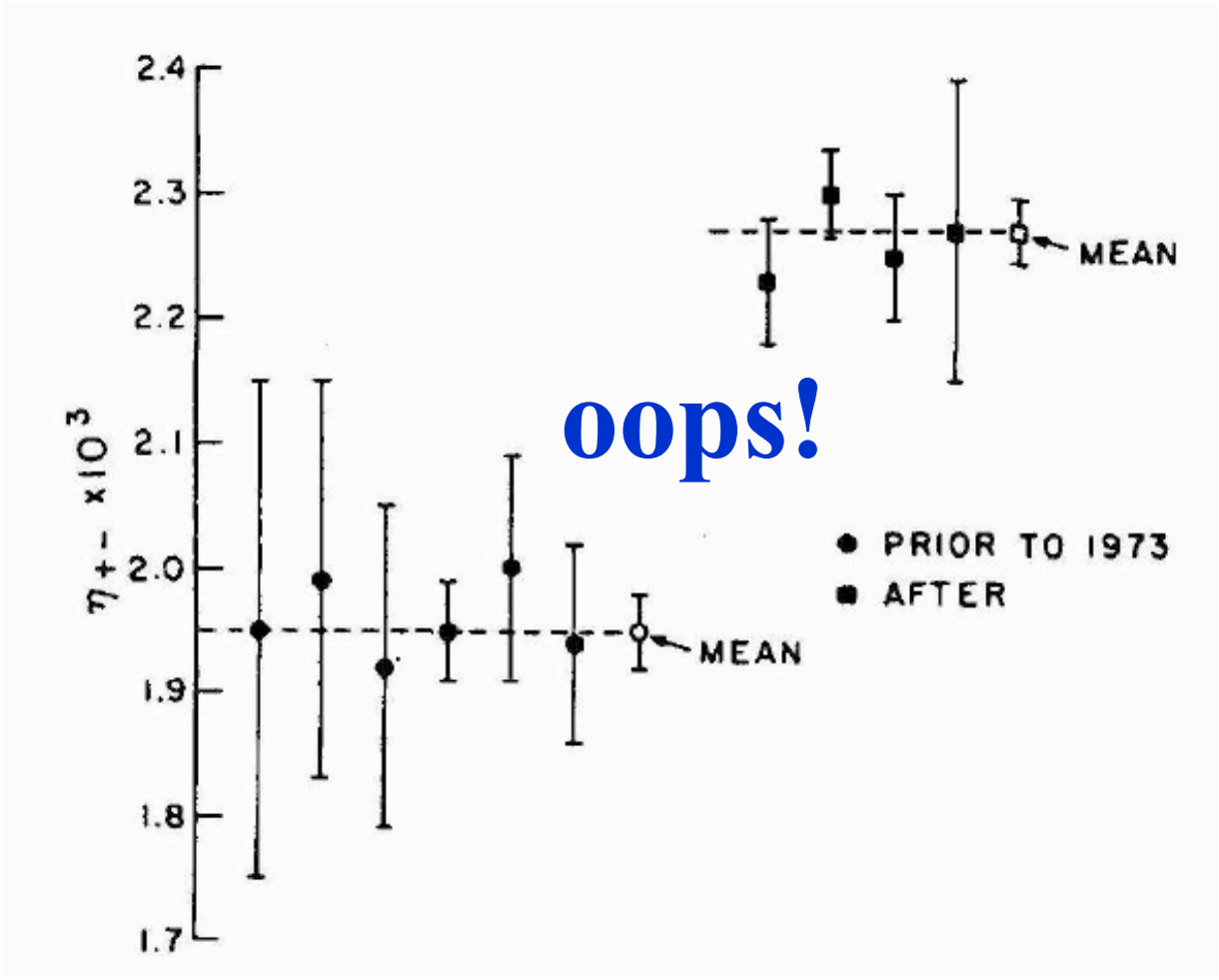
The key:

use standard formats



Excuse 2 – Wrong Papers!

Measurement of the $|\eta_{+-}|$ (a CP parameter) as a function of year.



A. Franklin, Am. J. Phys. 52, 786-793 (1984)

Excuse 3 – It Costs Too Much!

New York Times Headlines, **April 18, 2041**

"New membrane theory predicts strange effects at 7 TeV"

The cost of repeating the LHC in 2041 might well be $20 billion or more. It is unlikely to happen.

The cost of providing reliable, useable, open data *today* is estimated to be about 10% of typical HEP budgets…large, but not crazily so.

Excuse 4 – It's Not Fair!

…yes, it's not fair…to the ***taxpayers***!

We lobby governments to keep the tax money flowing to our experiments, but then we arrogantly insist that taxpayers are, in effect, unworthy of the data they paid for…

There is a workable compromise:

Scientists would have exclusive access to the data for a fixed period, say, 18 months, before the data become open.

Excuse 4 – It's Not Fair!

HIGH-ENERGY PHYSICS

Top quarks go it alone

The first, long-sought evidence for the production
has been reported from a sophisticated analysis o
collisions at the Tevatron.

www.physicstoday.org
physics today
May 2007

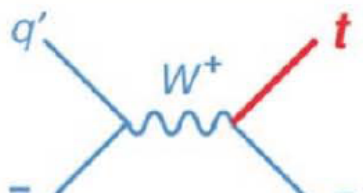
physics update
Suppleme
related to
can be fo
www.phy

SCIENTIFIC AMERICAN

BLACK HOLE BLOWBACK:
Building Galactic Clusters

MARCH 2007
WWW.SCIAM.COM

Unpaired top quarks. Weighing 200 times as much as a proton, the top quark is by far the heaviest elementary particle known. Because the strong nuclear force can't change a quark's flavor, it can produce quarks only in pairs with their antiquarks. The weak force can change flavors. But weak-interaction cross sections are so small that it's almost impossible



PHYSICS

Alone at the Top

CLOSER TO GOD: FERMILAB MAKES SOLO TOP QUARKS BY ALEXANDER HELLEMANS

The world's biggest accelerator, the physics, top quarks may emerge in collisions

The Citizen Scientist

All Science to All People!

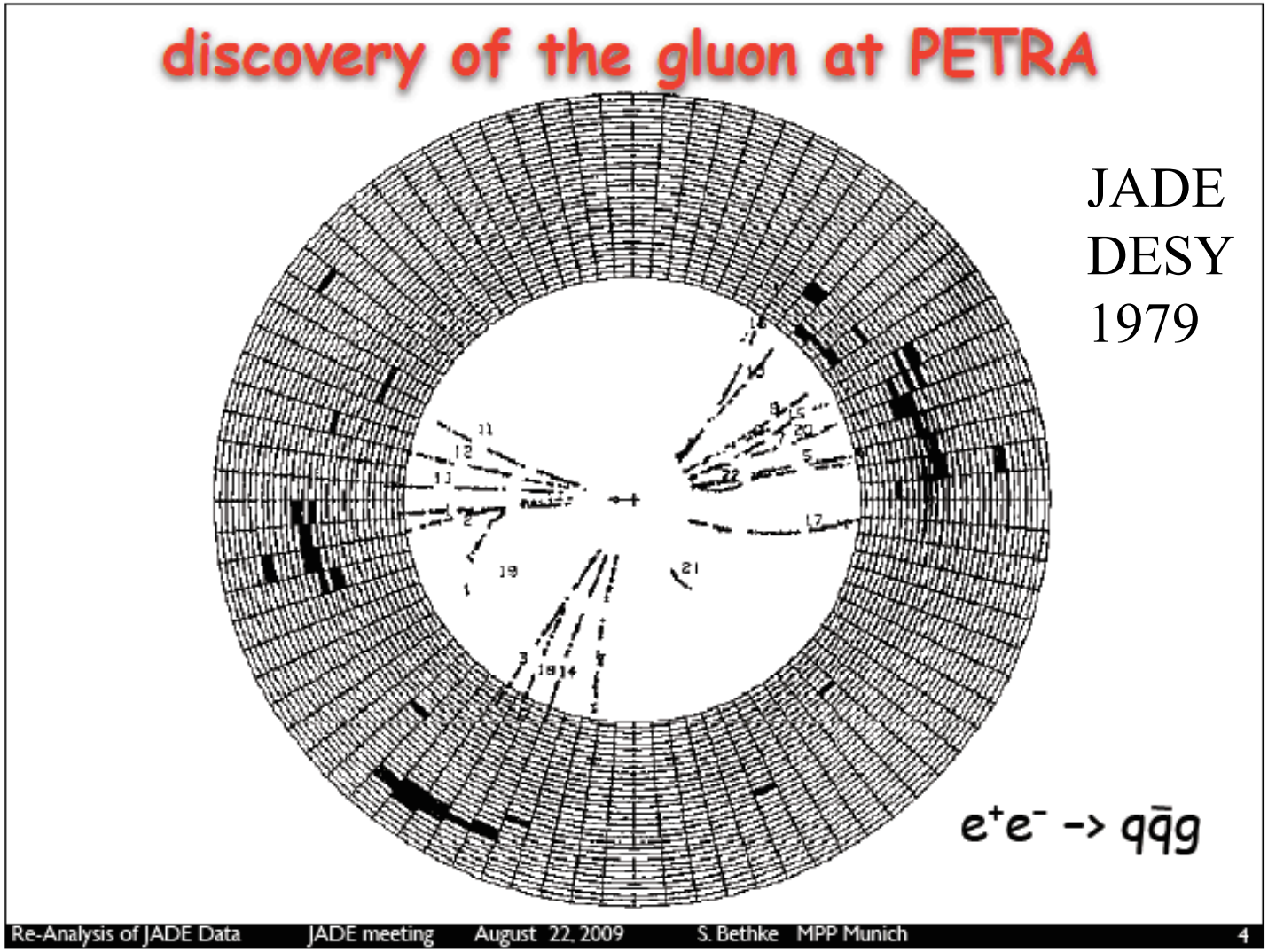
"why don't I feel the Earth move?"

With that question, my (then very young) daughter made a first small step towards becoming a citizen scientist.

A ***citizen scientist*** is

anyone who contributes to the corpus of scientific knowledge and understanding.

Citizen Science – JADE



Citizen Science – JADE

Publications based on resurrected JADE data (1997-2009)

7) Study of moments of event shapes and a determination of alpha(S) using e+ e- annihilation data from Jade.
Christoph Pahl (Munich, Max Planck Inst. & Munich, Tech. U.) , Siegfried Bethke, Stefan Kluth, Jochen Schieck, the JADE collaboration (Munich, Max Planck Inst.) . MPP-2008-135, May 8, 2009. 14pp. Eur.Phys.J.C60:181-196,2009, e-Print: arXiv:0810.2933 [hep-ex]

6) Determination of the Strong Coupling alpha(S) from hadronic Event Shapes and NNLO QCD predictions using JADE Data.
By JADE Collaboration (S. Bethke et al.). MPP-2008-131, Oct 2008. 9pp., Submitted to Eur.Phys.J.C. e-Print: arXiv:0810.1389 [hep-ex]

5) Measurement of the strong coupling alpha(s) from the four-jet rate in e+ e- annihilation using JADE data.
By JADE Collaboration (J. Schieck et al.). MPP-2006-161, 2006. 11pp. Eur.Phys.J.C48:3-13,2006, Erratum-ibid.C50:769,2007.
e-Print: arXiv:0707.0392 [hep-ex]

4) Measurement of the longitudinal and transverse cross-sections in e+ e- annihilation at s**(1/2) = 35-GeV - 44-GeV.
By JADE Collaboration (M. Blumenstengel et al.). MPI-PHE-2001-11, Jun 2001. 12pp., Phys.Lett.B517:37-46,2001.
e-Print: hep-ex/0106066

3) QCD analyses and determinations of alpha(s) in e+ e- annihilation at energies between 35-GeV and 189-GeV.
By JADE collaboration and OPAL Collaboration (P. Pfeifenschneider et al.). CERN-EP-99-175, Dec 1999. 49pp.
Eur.Phys.J.C17:19-51,2000. e-Print: hep-ex/0001055

2) C parameter and jet broadening at PETRA energies.
By JADE Collaboration (O. Biebel et al.). PITHA-98-21A, Mar 1999. 14pp. Phys.Lett.B459:326-334,1999., e-Print: hep-ex/9903009

1) A Study of event shapes and determinations of alpha-s using data of e+ e- annihilations at s**(1/2) = 22-GeV to 44-GeV.
By JADE Collaboration (P.A. Movilla Fernandez et al.). PITHA-97-27, Aug 1997. 36pp. , Eur.Phys.J.C1:461-478,1998.
e-Print: hep-ex/9708034

==========

3) Tests of analytical hadronisation models using event shape moments in {\epem} annihilation.
C. Pahl, S. Bethke, O. Biebel, S. Kluth, J. Schieck . MPP-2009-38, Apr 2009. 17pp. e-Print: arXiv:0904.0786 [hep-ex]

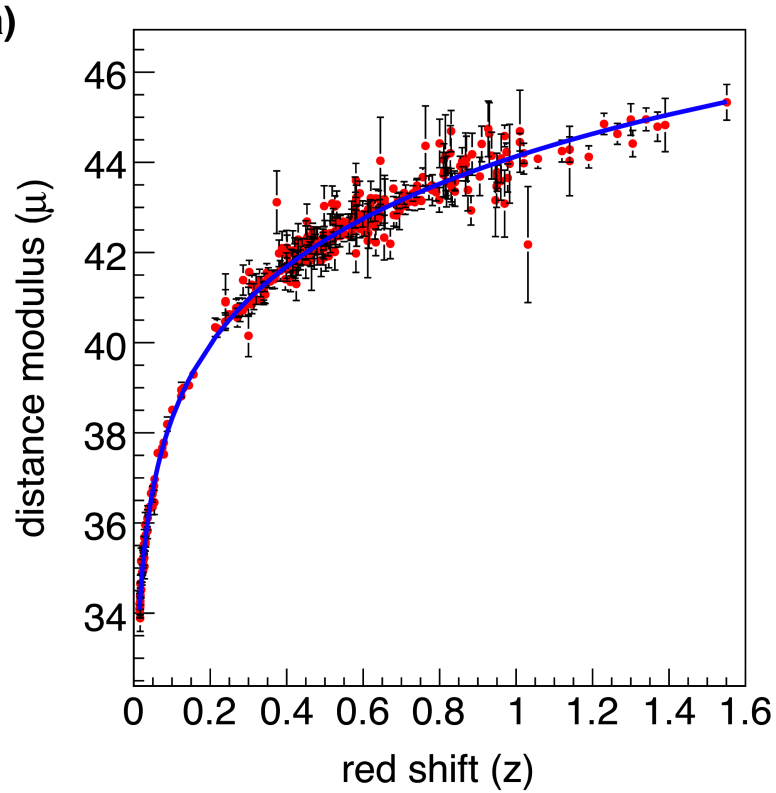
2) Tests of power corrections for event shapes in e+ e- annihilation. P.A. Movilla Fernandez, S. Bethke, O. Biebel, S. Kluth (Munich, Max Planck Inst.) . MPI-PH-2001-005, May 2001. 27pp., Eur.Phys.J.C22:1-15,2001. e-Print: hep-ex/0105059

1) A Measurement of the QCD color factors using event shape distributions at s**(1/2) = 14-GeV to 189-GeV.
S. Kluth, P.A. Movilla Fernandez, S. Bethke, C. Pahl, P. Pfeifenschneider (Munich, Max Planck Inst.) . MPI-PHE-2000-19, Dec 2000. 25pp.
Eur.Phys.J.C21:199-210,2001. e-Print: hep-ex/0012044

Re-Analysis of JADE Data JADE meeting August 22, 2009 S. Bethke MPP Munich 9

Citizen Science – Another Example

Can varying-G models explain the supernovae data? (Rutger Dungan)



Varying-G cosmology with type Ia supernovae

Rutger Dungan and Harrison B. Prosper
Department of Physics, Florida State University, Tallahassee, Florida 32306

(Received 25 September 2009; accepted 17 August 2010)

The observation that type Ia supernovae (SNe Ia) are fainter than expected give
led to the conclusion that the expansion of the universe is accelerating. T
hypothesis is that this acceleration is caused by a cosmological constant or sor
that pervades the universe. We explore what the supernovae data tell us abou
answering the question: Can these data be explained with a model in which th
varies on a cosmic timescale? We conclude that they can and find that the super
insufficient to distinguish between a model with a cosmological constant and on
However, the varying-G models are not viable when other data are taken into account. The topic is
an ideal one for undergraduate physics majors. © *2011 American Association of Physics Teachers.*
[DOI: 10.1119/1.3486585]

I. INTRODUCTION

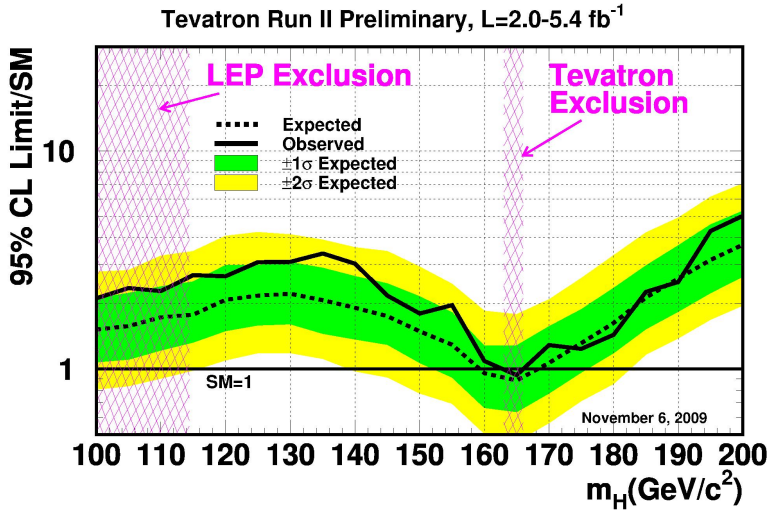
Physical cosmology is concerned with the large scale

$$ds^2 = c^2dt^2 - a^2(t)\left[\frac{dr^2}{1-Kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right], \quad (2)$$

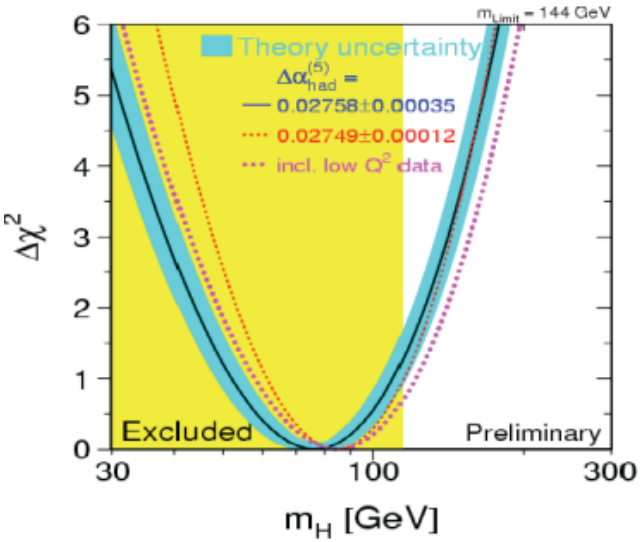
Am. J. Phys. 79, 62 (2011)

Citizen Science – Missed Opportunity

If a SM Higgs exists, we know a lot about it:



$p(s|m)$



$p(m)$

But because HEP data are closed, we cannot do much with these plots.

$p(s)$

Towards Open Data

From Secrecy…

In 1609, Galileo turned his telescope to Saturn and encrypted his discovery in an anagram:

s m a i s m r m i l m e p o e t a l e u m i b u n e n u g t t a u i r a s

which when unscrambled read:

Altissimum planetam tergeminum observavi

"I have observed the highest planet to have a triple form."

…to Openness

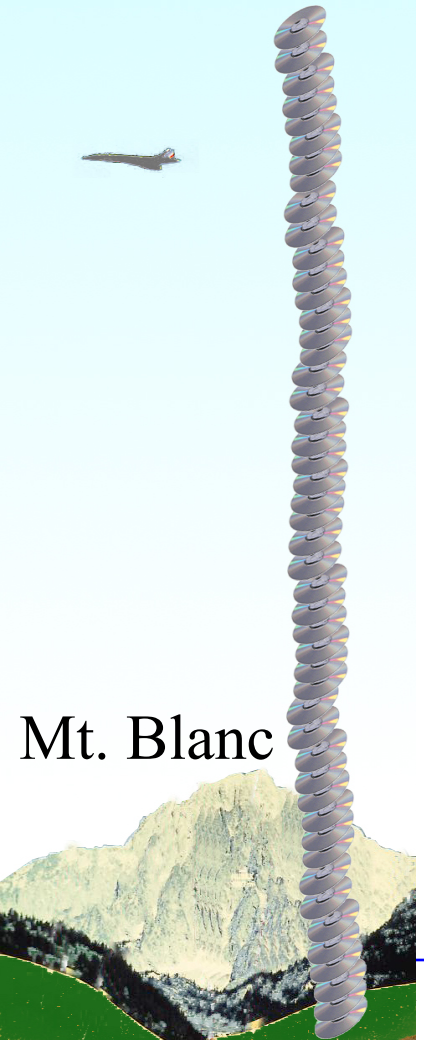
CERN Convention (1953):

"…the results of its experimental and theoretical work shall be published or otherwise made ***generally available***"

We have had open access to *published* results for decades.

It is now time to interpret the CERN statement more broadly, to imply open access to *data* also.

A 20 km stack of cds = 1 year of LHC data



The Data Tower

Data Type	Per Event	Per Year
Raw	1.6 MB	3200 TB
Reconstructed	1.0 MB	2000 TB
Physics	0.1 MB	200 TB

…per LHC experiment!

However, today there are no technical roadblocks to making these data open.

A Modest Proposal for 2011

1. Release all CDF and Dzero internal notes to INSPIRE

2. Make *physics* data associated with publications open.

   - Release the likelihood functions $p(D \mid \theta, \omega)$ of all analyses in RooStats workspaces.

   - Have CMS and ATLAS Collaborations officially endorse *open source* fast simulations of their respective detectors. An excellent candidate is Delphes (http://www.fynu.ucl.ac.be/users/s.ovyn/Delphes/index.html). Another is TurboSim.

TurBoSim

Developed at Dzero by Bruce Knuteson in the 1990s.

TurboSim

A fast, self-tuning, detector simulation

$p(D \mid I)$

Basic idea:

Abstract the map

Detector($D \mid O$)

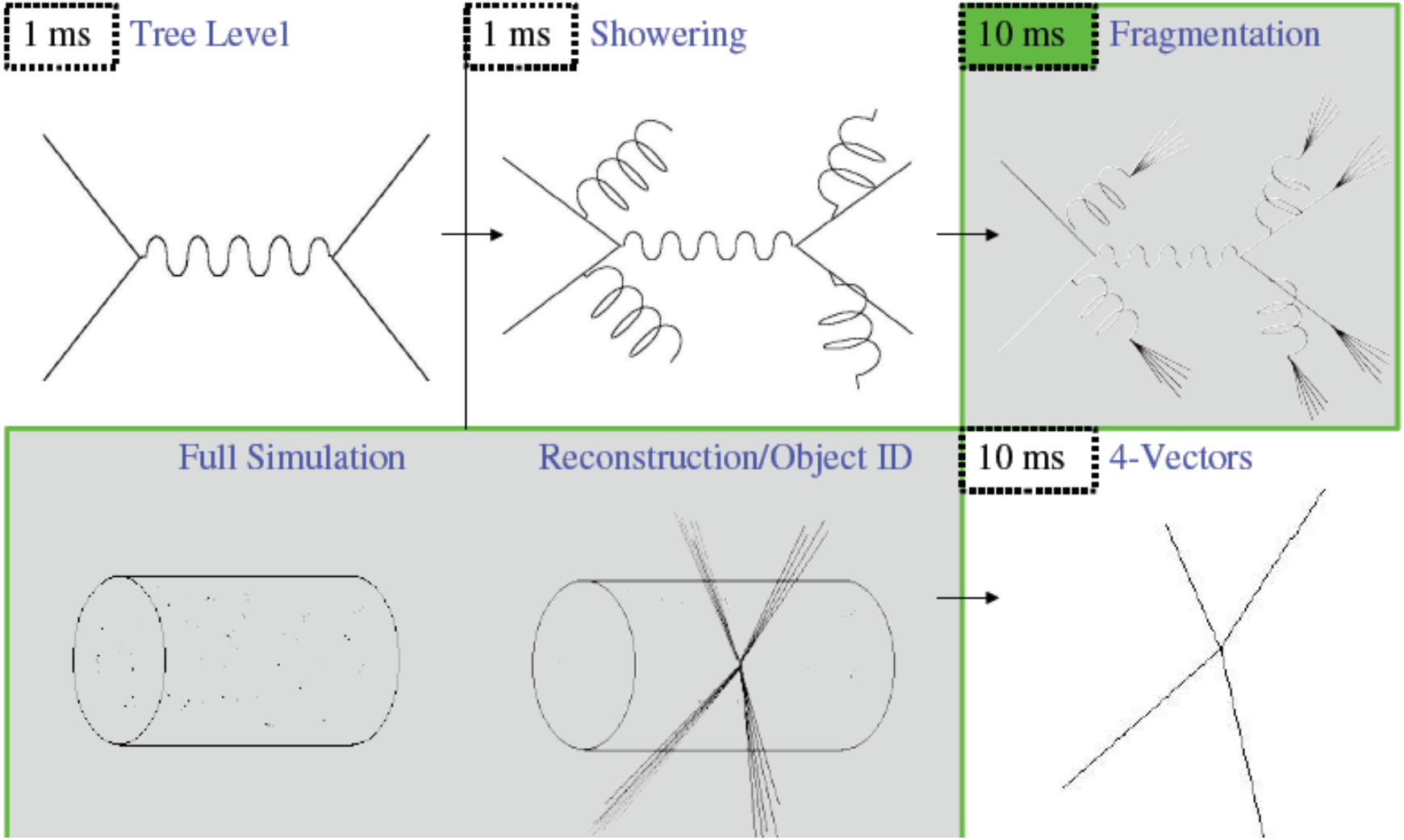
Bruce Knuteson

David Friend

Khaldoun Makhoul

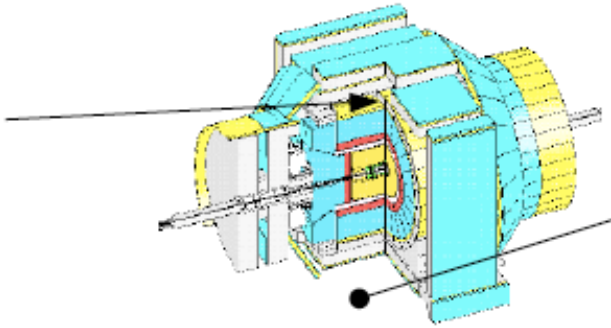
from fully simulated events and represent as a lookup table.

Simulation with TurboSim

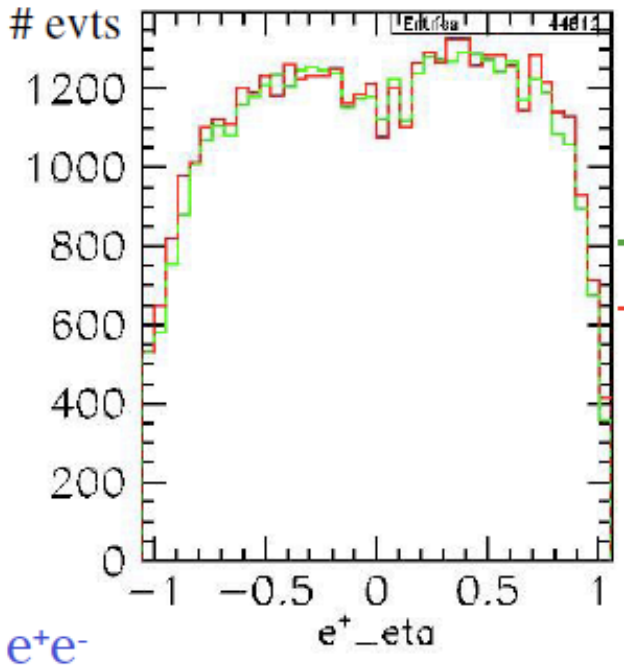


Comparison Plots: CDF

Crack in EM Calorimeter at eta =0



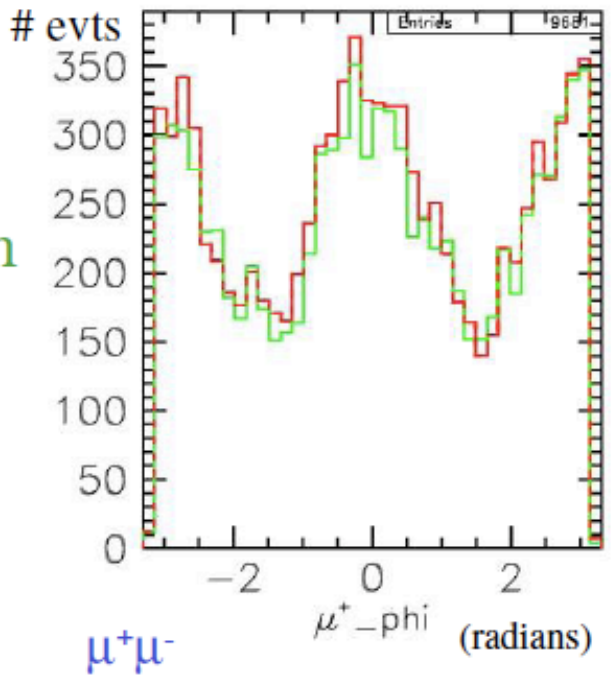
Muon coverage varies with phi



e⁺e⁻

TurboSim

CdfSim



μ⁺μ⁻

Summary

- Open data will enable more science.
- Some of this science will be bad, some will be excellent, and now and then, some will be extraordinary.
- The move towards open data is gaining pace.
- Rather than wait for an open data policy to be imposed from on high (by the funding agencies) we should take the initiative and embrace open data…now!