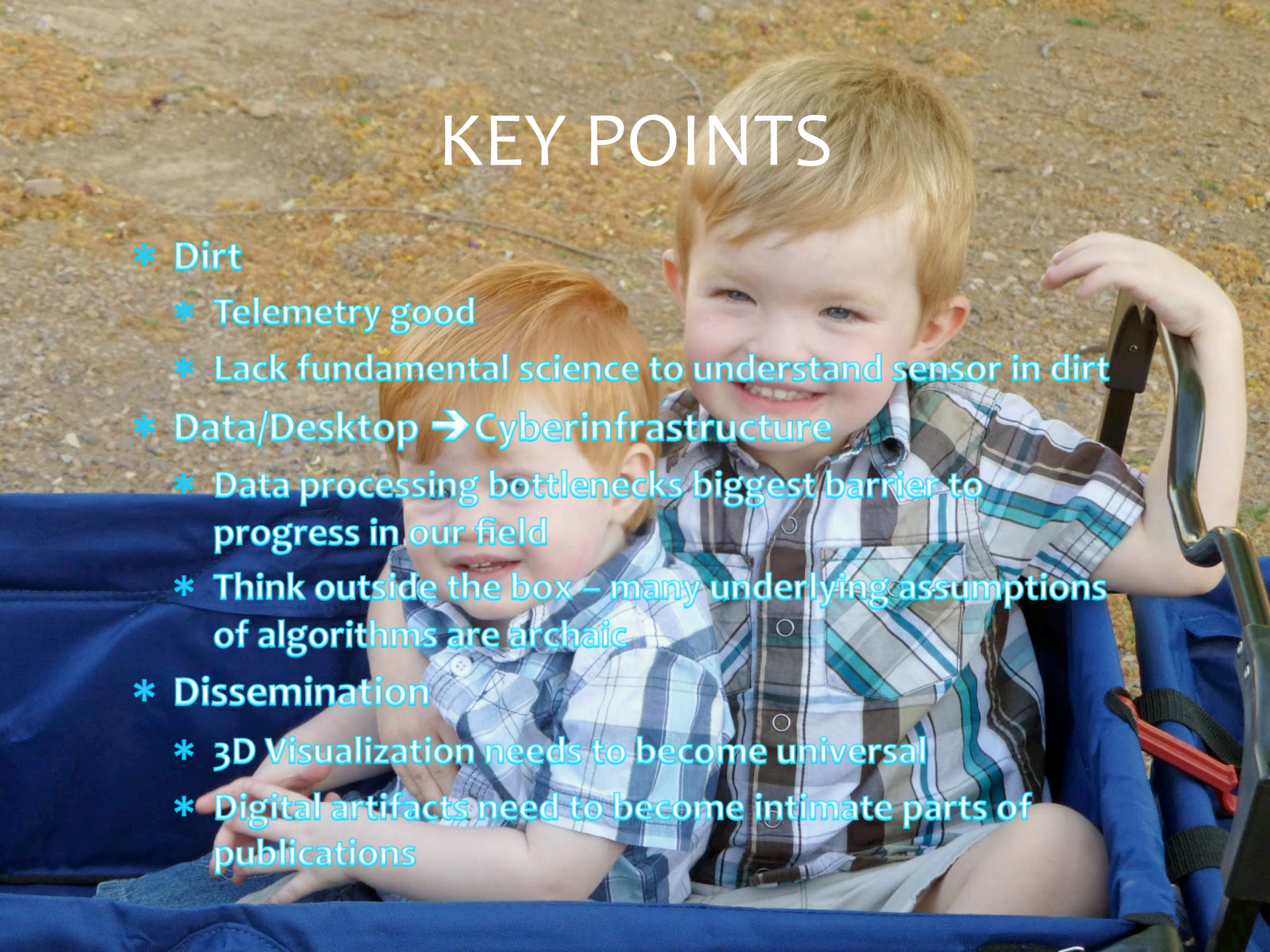
Lessons Learned from 24 Years of Collecting and Processing Passive Seismic Array Data

Gary L. Pavlis
Department of Geological Sciences
Indiana University



KEY POINTS

* Dirt
  * Telemetry good
  * Lack fundamental science to understand sensor in dirt
* Data/Desktop → Cyberinfrastructure
  * Data processing bottlenecks biggest barrier to progress in our field
  * Think outside the box – many underlying assumptions of algorithms are archaic
* Dissemination
  * 3D Visualization needs to become universal
  * Digital artifacts need to become intimate parts of publications

My History

- Pinyon HF array – 1990
- Pinyon BB array – 1991
- KNET- 1991+
- Kazakh array – 1992
- Geyokcha, Turkmensitan – 1993
- Wabash Valley – 1995
- Lodore 1995-1996
- Tien Shan 1999-2001
- Bolivar 2002-2004
- STEEP 2003-2008
- OIINK 2011-present
- Homestake 3D array 2015

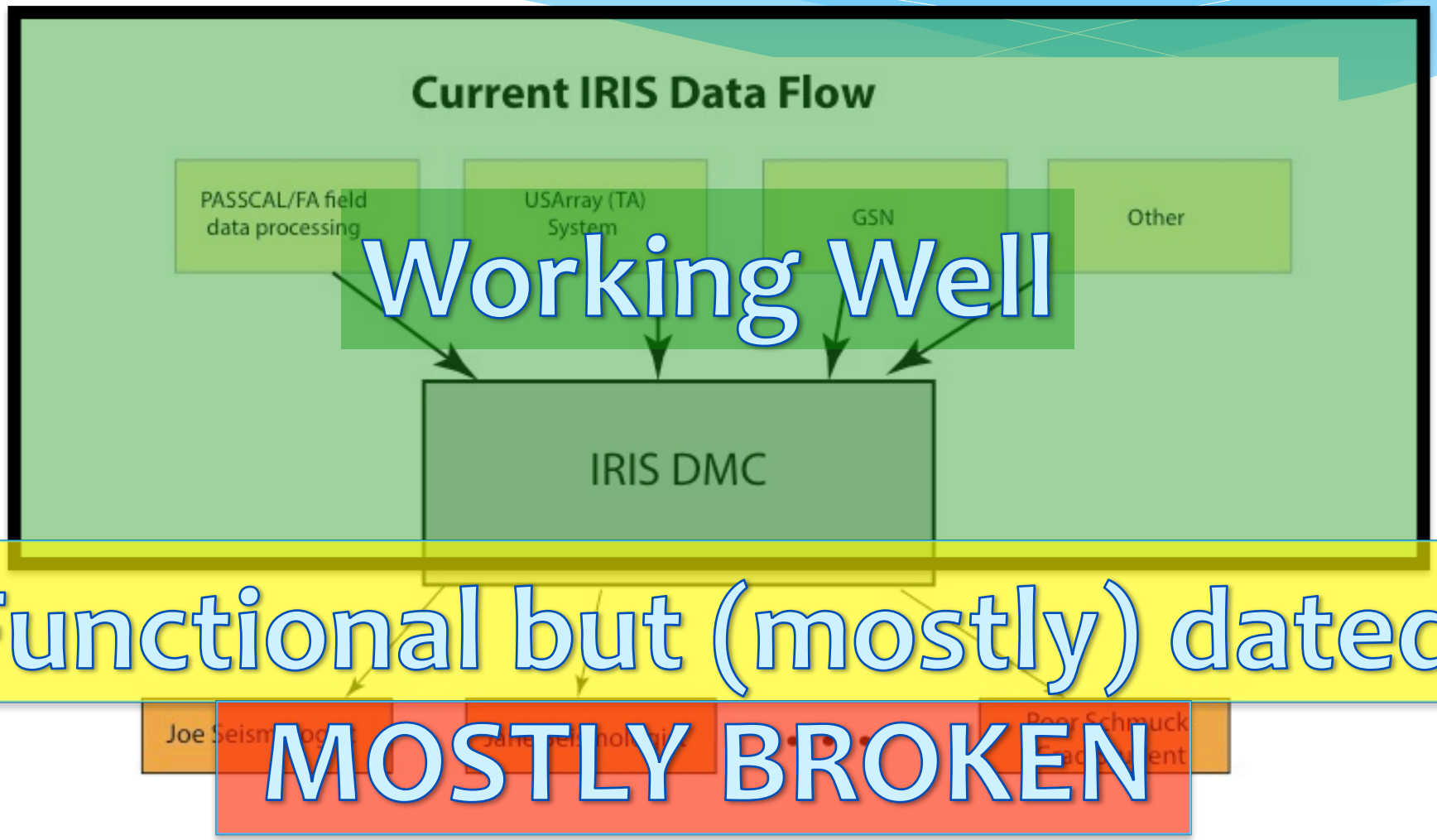
Data/Desktop

* Topics are inseparable -> Cyberinfrastructure
* Changes in computing in 24 years of my title
  * Desktop speed increased by more than 1000
  * Supercomputers today max ~10^{13} flops
  * HPC storage systems ~10^{16} bytes
* Important quote: David Hale at first USArray Data Processing workshop: "Factors of 1000 are hard on assumptions"

Corollaries of David Hale quote

* Computational expedience is a hidden assumption of many methods we use (choice made decades ago lost in evolutionary development)
* High performance computing then is desktop now
* We all need to “think outside the box” and seek out these hidden assumptions

Example of the issues: IRIS Data Model



Why can you assert data processing infrastructure is broken?

* USArray Data Processing course instructor yearly since 2009
  * We've collected best use cases we can find
  * I know what you and your students do
* If you process data with a commercial CMP processing system like ProMAX you understand the difference

The difference is like this

What we use

Commercial CMP systems

Why

What we use

* Irregular maintenance
* Clashing conceptual models for data handling
* Research code
* Full of archaic stuff like SAC and duct tape home brew shell scripts
* Poverty in comparison to any major oil and gas company

Commercial CMP systems

* Well maintained
* Time is money
* Always kept state-of-the-art
* Megacorporations with deep pockets and budgets bigger than many countries

A Key Assertion

* The Cyberinfrastructure of today is completely analogous to the state of seismic instrumentation in 1980 pre-IRIS
  * Balkanized
  * Huge energy wasted by duplication of effort
  * Largest bottleneck to progress in the field today

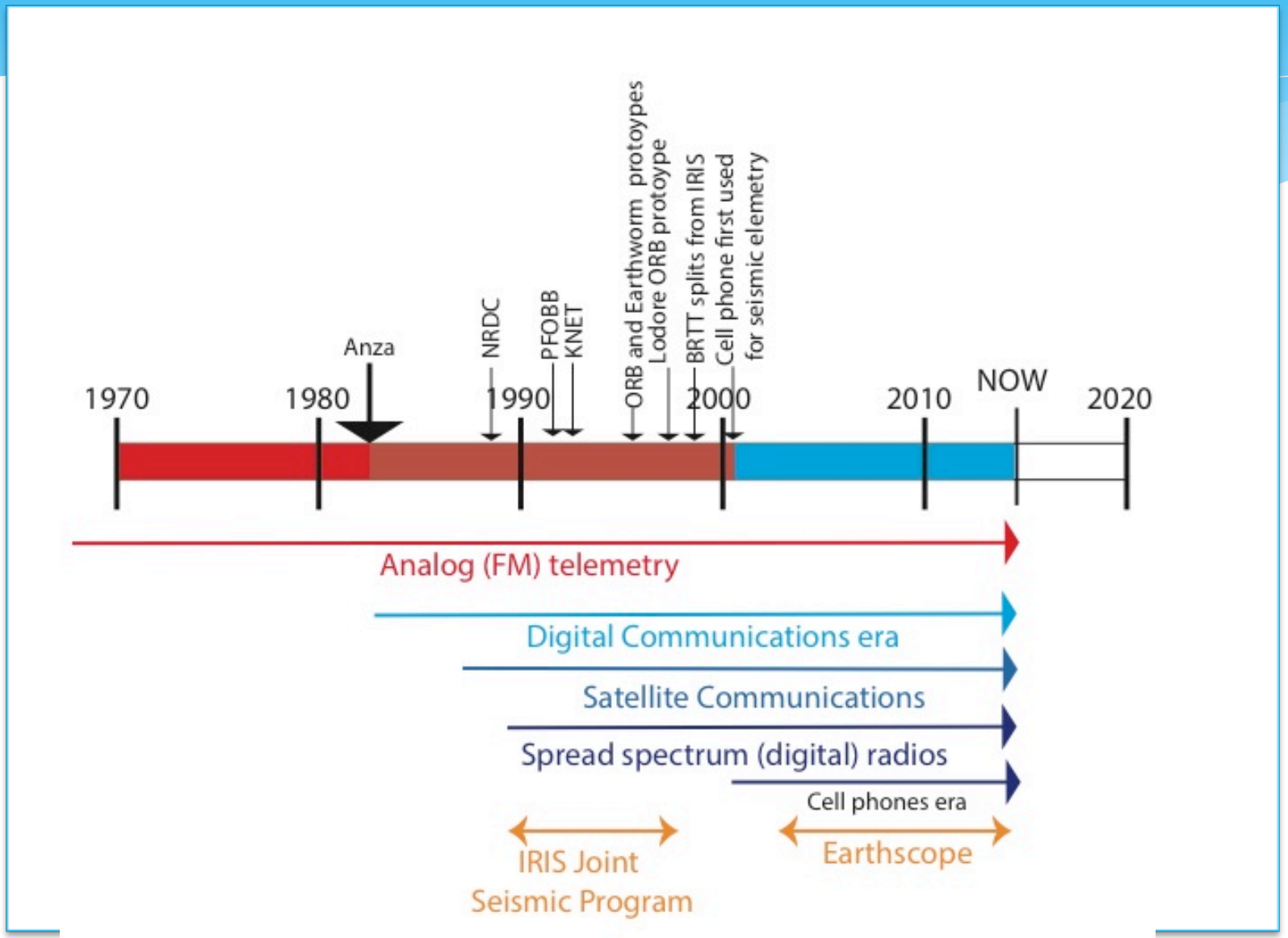
What we need: A working data processing framework

* Efficient data flow system
  * Abstracts and handle multiple common seismological objects
  * Passes data objects through system without intermediate disk files
* System for clean handling of auxiliary data
  * "Metadata" (What we used to call trace headers)
  * Processing algorithm input parameters (Fill in forms for parameters)
* System that knows something about parallel computing
* Clean API
  * Maximize community developments
  * API to packages people use (Matlab, Antelope, ObsPy, and (yes) SAC)
* (Maybe) Simplified "flow builder" ala Promax

How to do this?

* We need to do something about this and stop just talking about it
* Earthcube will not be solving this problem for us
* Form a technical working group to develop a development plan
* Put in at least some seed money to get this moving

DIRT 1: Telemetry
Brief, IRIS-centric History Lesson



Key Points on History

* IRIS Joint Seismic Program role in TA performance needs to be remembered
* Seismology has been a leader in digital sensor network developments
* We are opportunists that adapt to changing technologies well

Plus and Minus of Telemetry compare to standalone (PASSCAL) mode

Positive

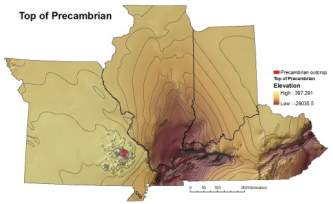
* Data quality
* Data recovery
* Faster data turnaround
* Can make experiment cheaper

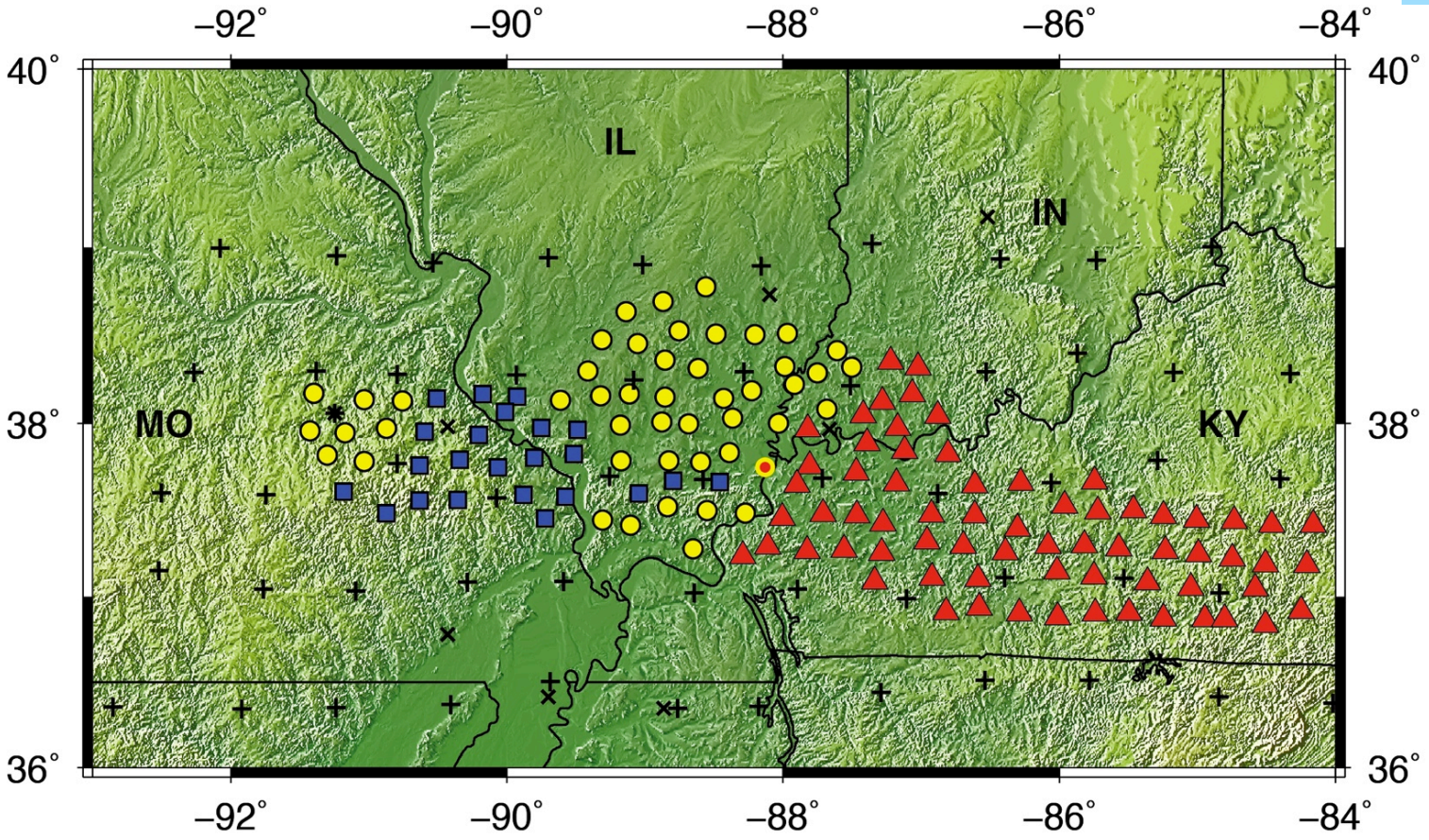
Negative

* Longer station install times
* Greater complexity
  * Operations
  * Data handling
* PI and team need more background education

OZARKS | ILLINOIS | INDIANA | KENTUCKY

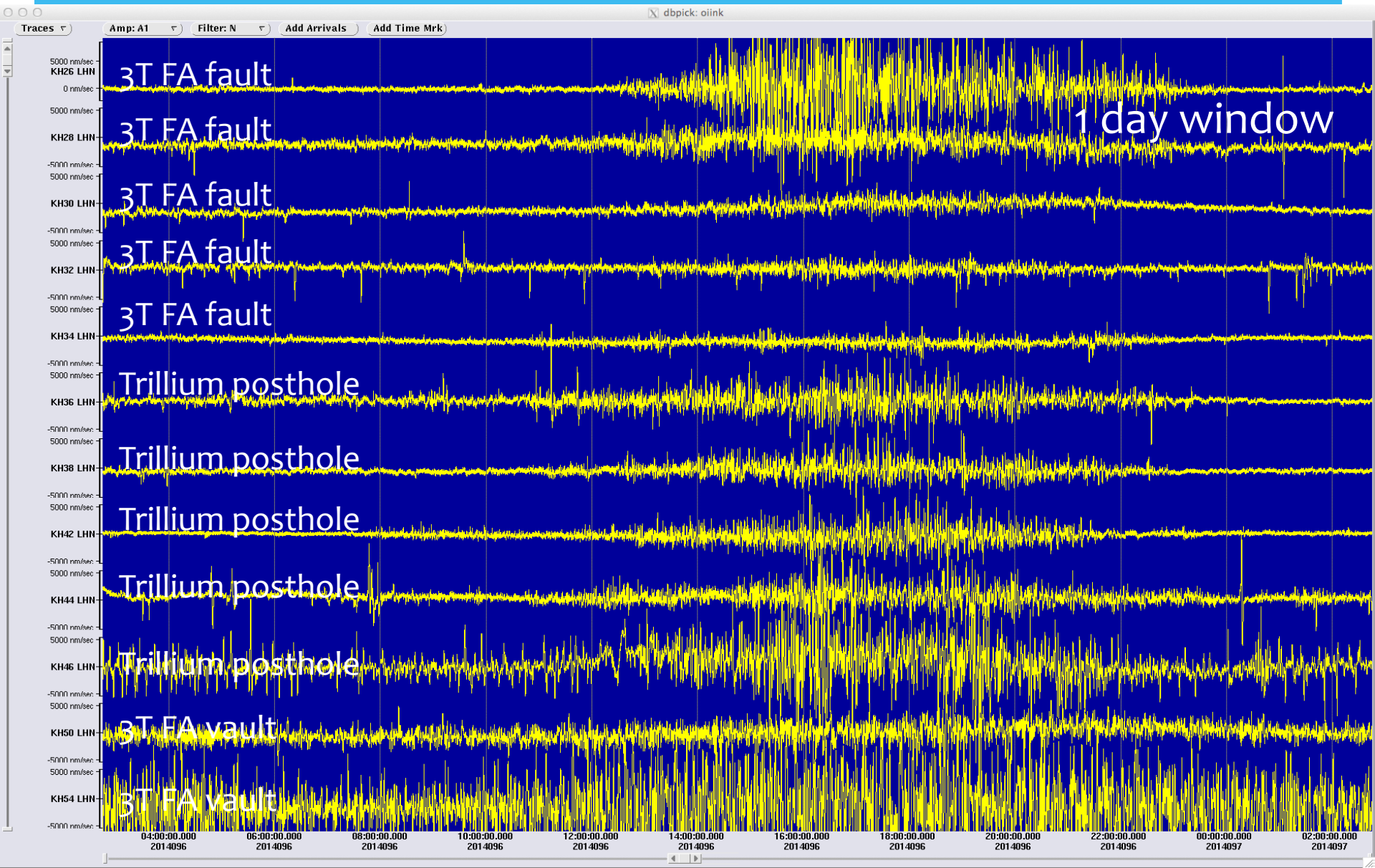
unraveling the secrets of North America's continental interior



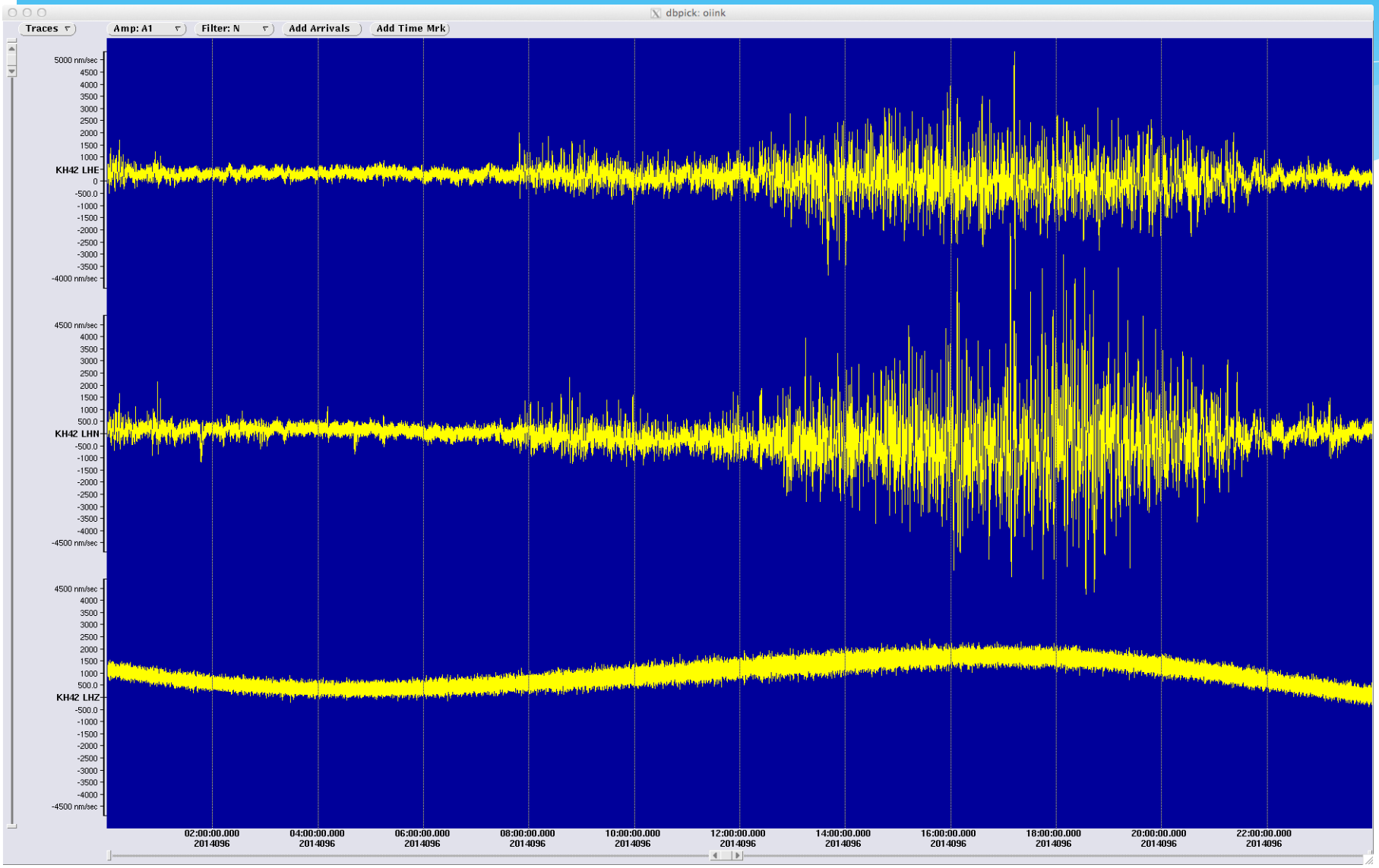


* 51-55 of 70 stations use cell phone telemetry at different phases
  * Data recovery > 99% on telemetry stations
  * Standalones around 97% recovery
  * "Good" data MUCH higher with telemetry
    * 20%+ Guralp 3T failure rates exaggerated difference
    * Outside failures, mass position problems rare with telemetry – common otherwise
* Saved us A TON of money
  * HUGE to us because USArray paid the cell bills
  * Cost analysis shows we would still have saved money had we paid the cell bill
  * Especially cost effective if you count difference in lost data

Dirt 2: BB sensors in Dirt



Fact: LP noise is dominated by tilt



This is an old, unsolved problem

My (now 30 year old) daughter

Some different approaches to BB sensor deployments

Concrete Pad – sensor wall separated from pad

FA Vaults: double walled sensor in dirt

What we sort of know empirically

* Nothing beats a rock site in a deep mine for LP noise
* Cement a concrete pad to rock and the details matter little
* In dirt results are wildly different and the devil is in the details
  * FA vaults, postholes, and direct burial in a plastic bag are equivalent
  * Massive containers in dirt like TA vaults help at 10 db scale or more
  * Older work demonstrated separation of concrete pad from walls of enclosure reduces LP noise – somewhat forgotten

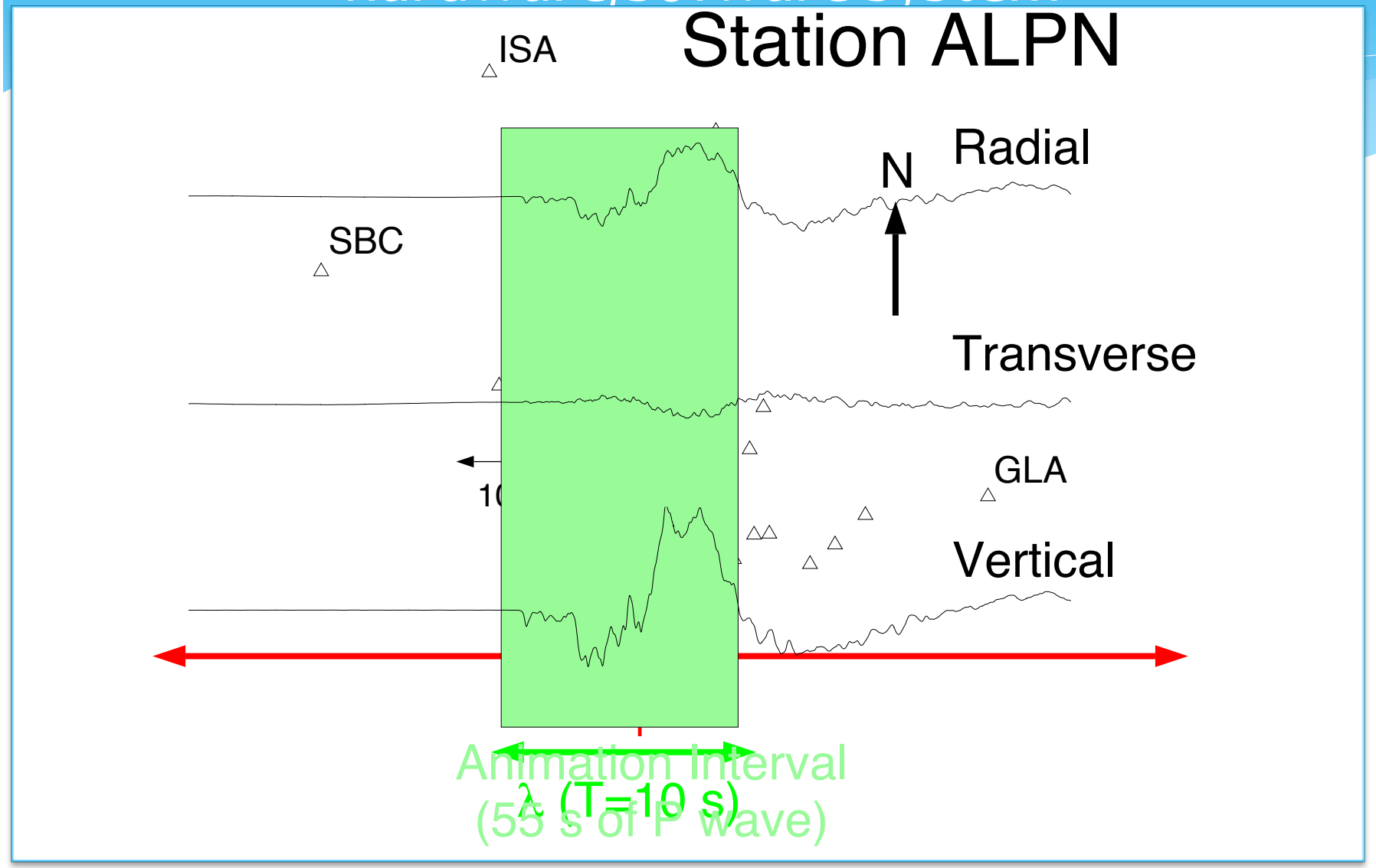
Why does this happen?

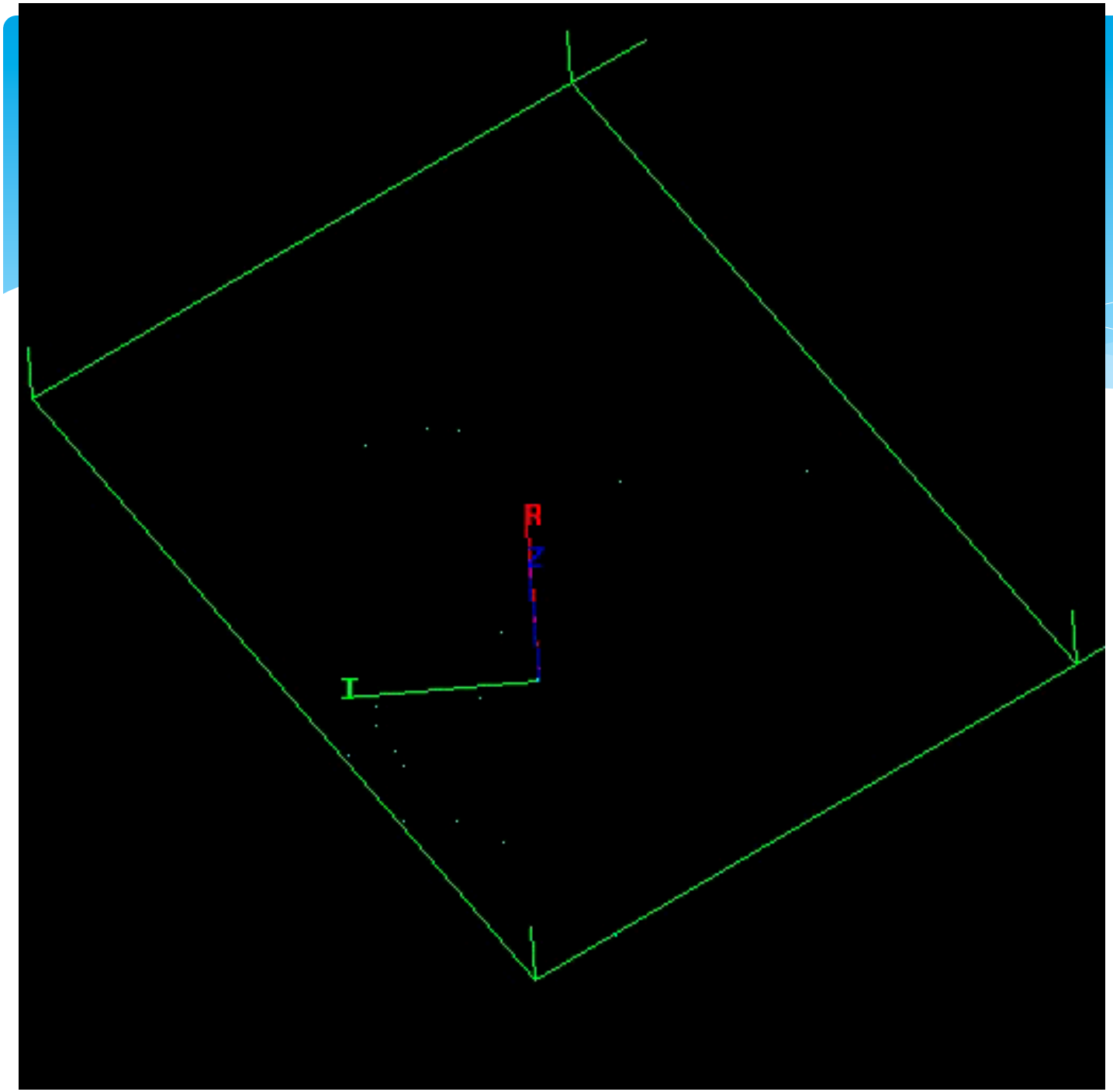
* What we know
  * Soil is unstable (it is why you can put a shovel in it)
  * Lots of processes tilt a sensor in the recording band of bb sensors
* What we don't know is the relative importance of
  * Thermal stress interaction with soil heterogeneity
  * Atmospheric pressure fluctuation interactions with soil heterogeneity
  * Wetting/drying induced stresses
  * Settling
  * Other processes?

Dissemination – 3D/4D visualization needs to become the norm

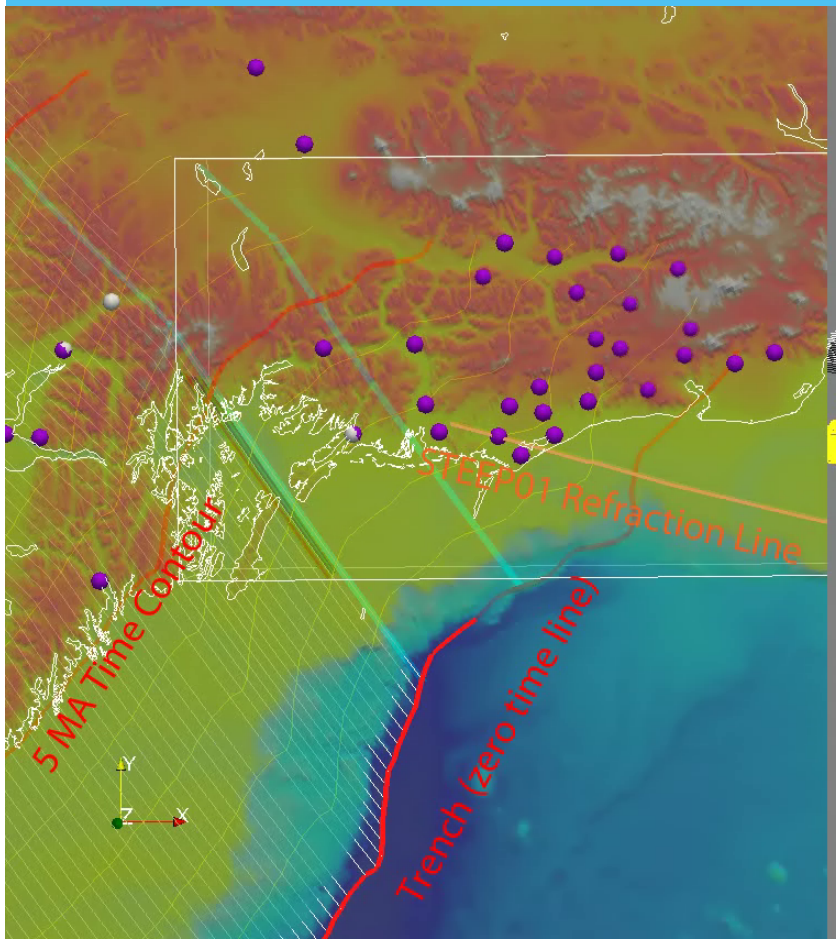
* Understanding 3D problems is a key element to advancing many questions in our science
  * Modeling 3D processes becoming common
  * Literature is full of misconceptions from 2D thinking applied to 4D problems
* 3D vis capabilities possible today on any reasonable hardware
* Publications are now close to 100% electronic

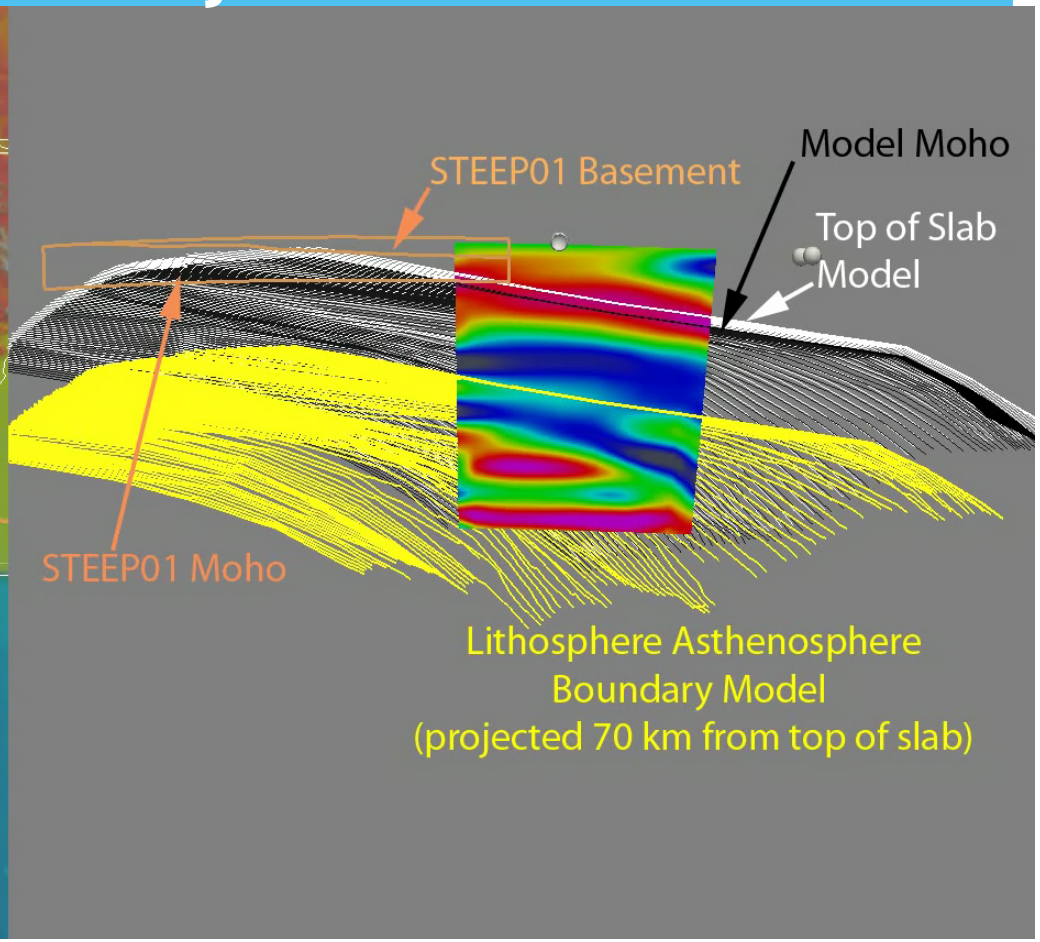
Late 1990s 3D visualization - $50,000 hardware/sofware system




R
I

3D Vis circa 2014- $1000 hardware/ software system



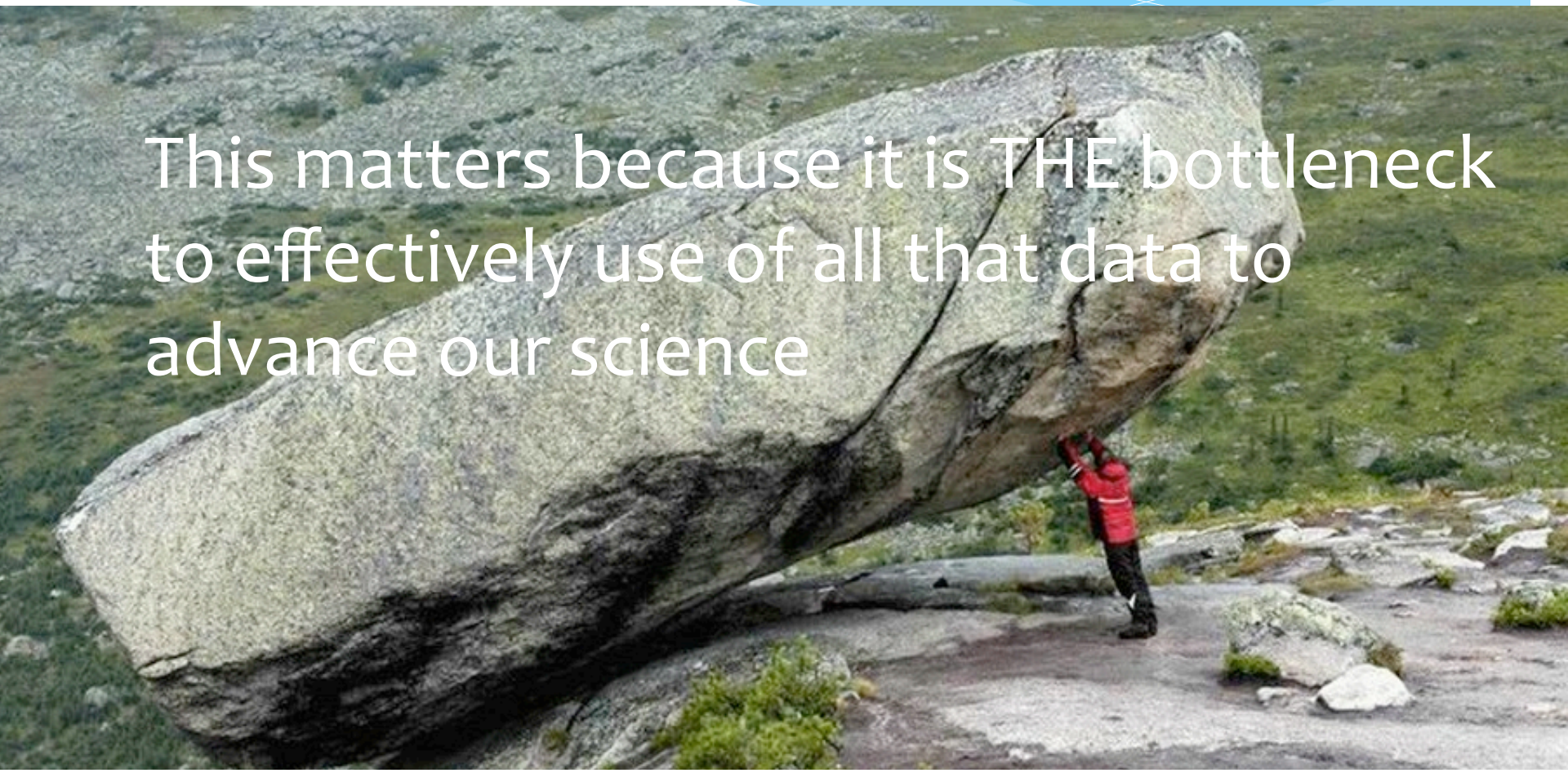


Bauer et al. (2014)

Summary

* Dirt
  * Telemetry – use it is you can
  * Basic research needed to understand processes that create LP tilt noise on broadbands
* Data/Desktop ➔ Cyberinfrastructure
  * Data processing infrastructure is THE biggest barrier to progress in our field today
  * Think outside the box – implicit assumptions of almost every processing algorithm we use should be questioned
* Dissemination
  * 3D visualization needs to become universal
  * Digital artifacts need to become part of all publications

The Way I View Too Many Data Processing Tasks Today



http://funnothingmore.blogspot.com/2012/10/impossible-photo-assignment.html

Tripod – Selwyn Sacks