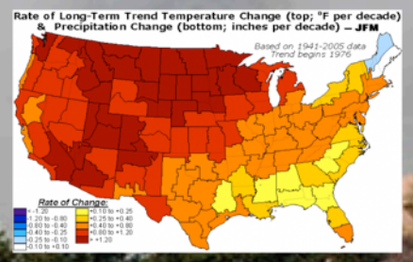
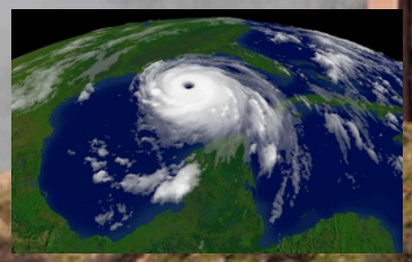
NCPP QED-2013
Quantitative Evaluation of Downscaling

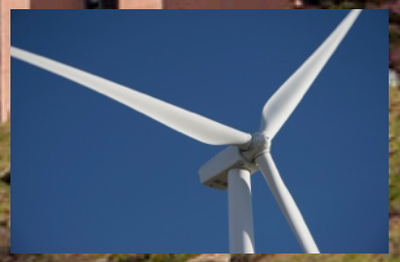
ClimateTranslator

Caspar Ammann, Jonathan Vigh
and NCPP Core / Tech Teams

air • planet • people



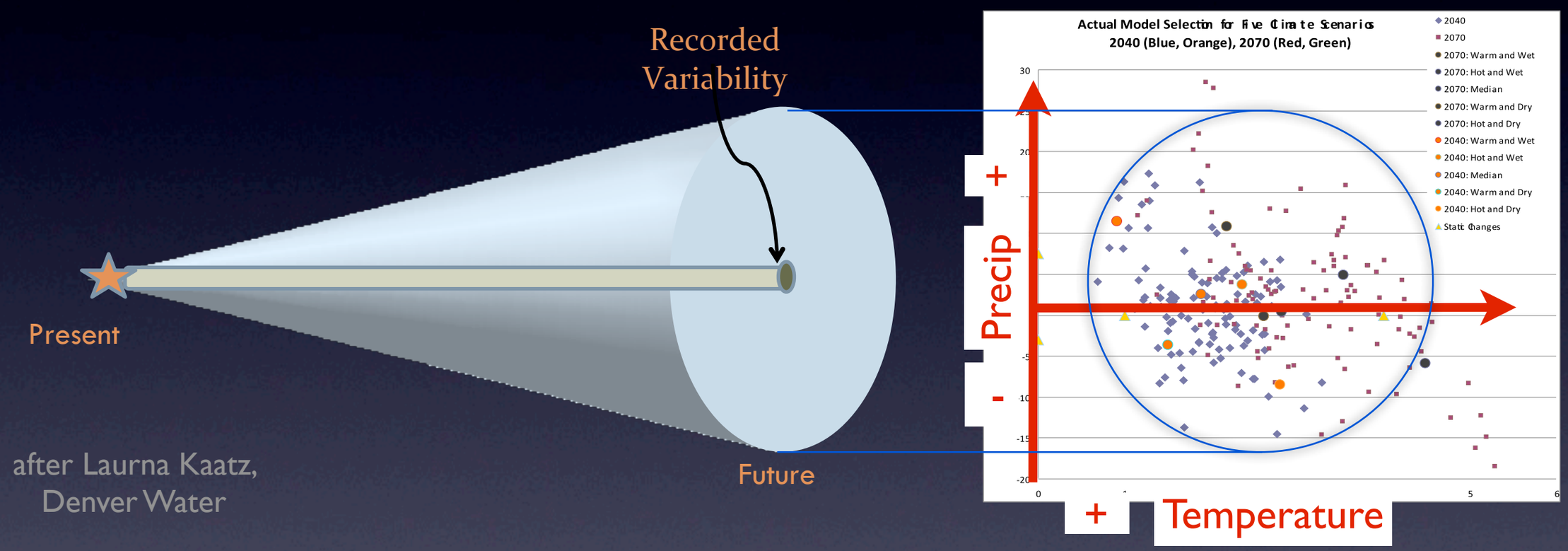




Workshop, August 12-16 2013

National Center for Atmospheric Research

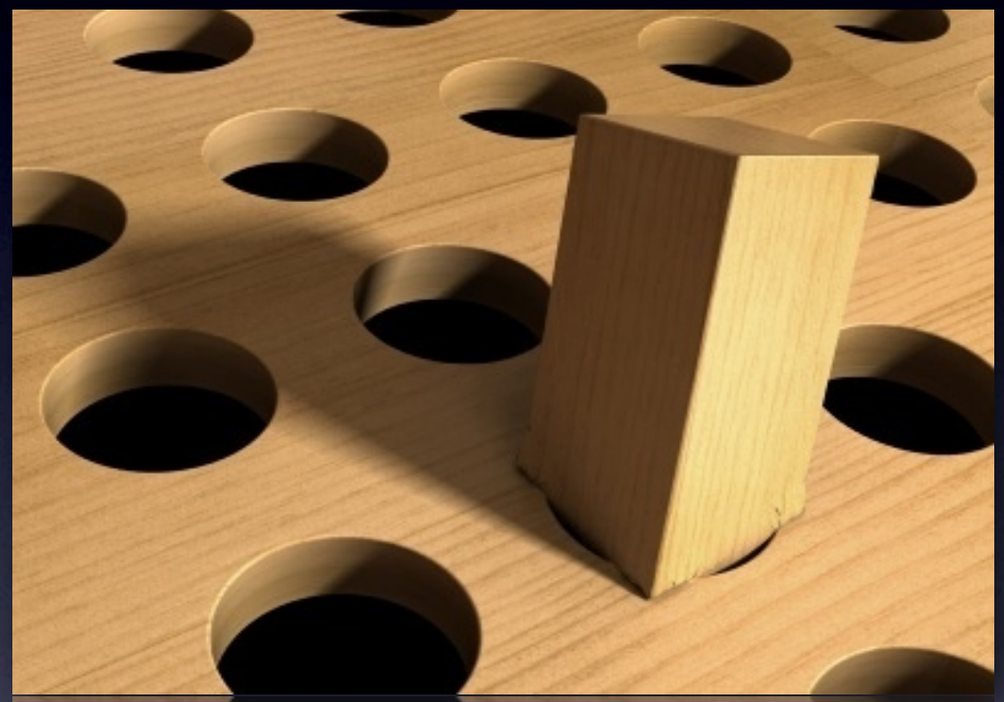
The "Practitioners Dilemma"



after Laurna Kaatz, Denver Water

Need for a ClimateTranslator

Need for a ClimateTranslator



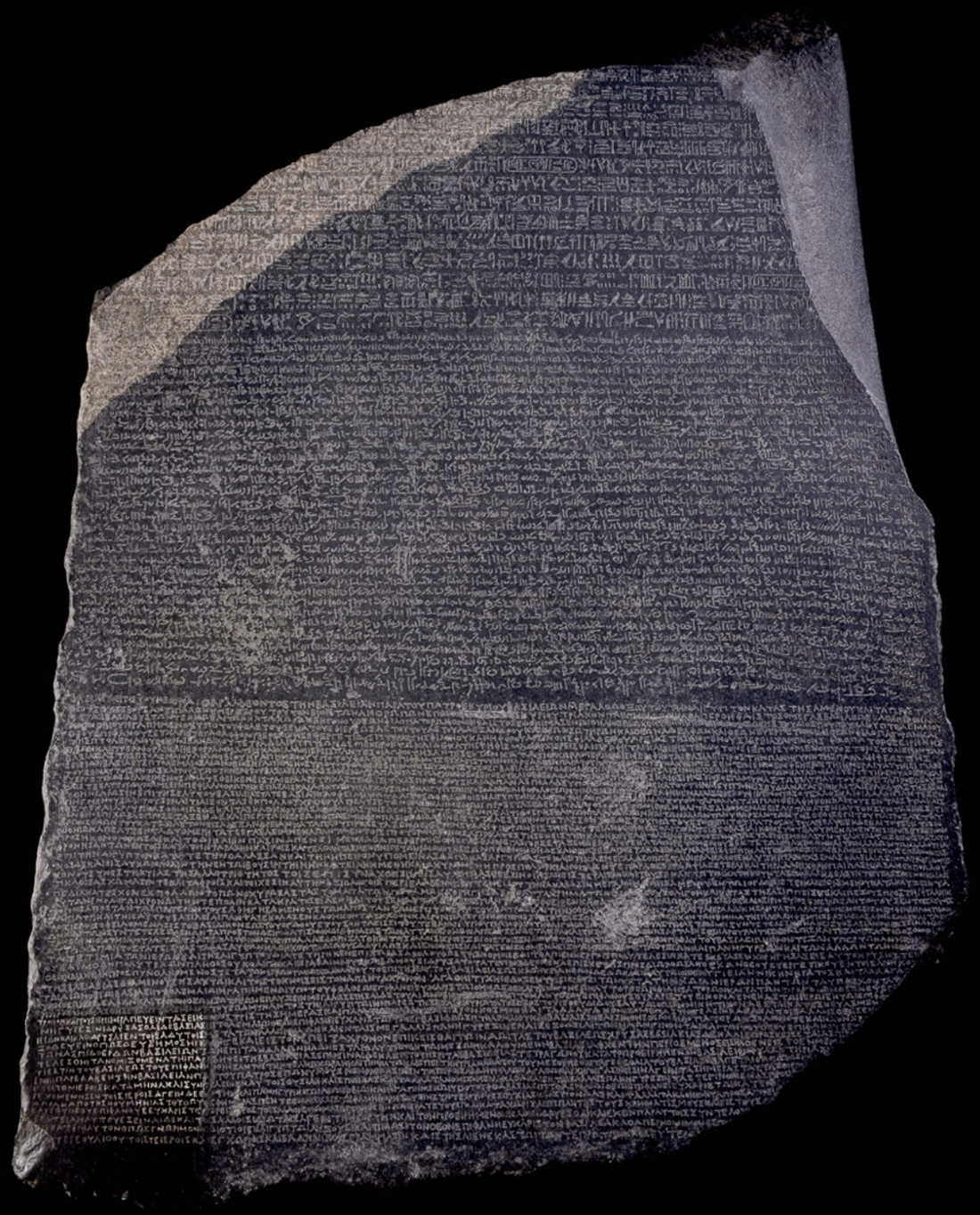
- **Data:** access, use (format, index, resolution)
  Which data to use? How to read it?
  Where does it come from?

- **Evaluation:** Quality Control Inter-comparison, data content info
  How good is data? Production assumptions?
  What are the uncertainties?

- **Translation of Scientific Knowledge** for exploration of impacts of change, guidance of use
  What does it mean? What is likely, what possible?
  Change in context?

- **Community of Practice**
  Collaboratively develops data requirements, handling of scenarios

Need for a ClimateTranslator



- **Data:** access, use (format, index, resolution)
  Which data to use? How to read it?
  Where does it come from?

- **Evaluation:** Quality Control
  Inter-comparison, data content info
  How good is data? Production assumptions?
  What are the uncertainties?

- **Translation of Scientific Knowledge** for exploration of impacts of change, guidance of use
  What does it mean? What is likely, what possible?
  Change in context?

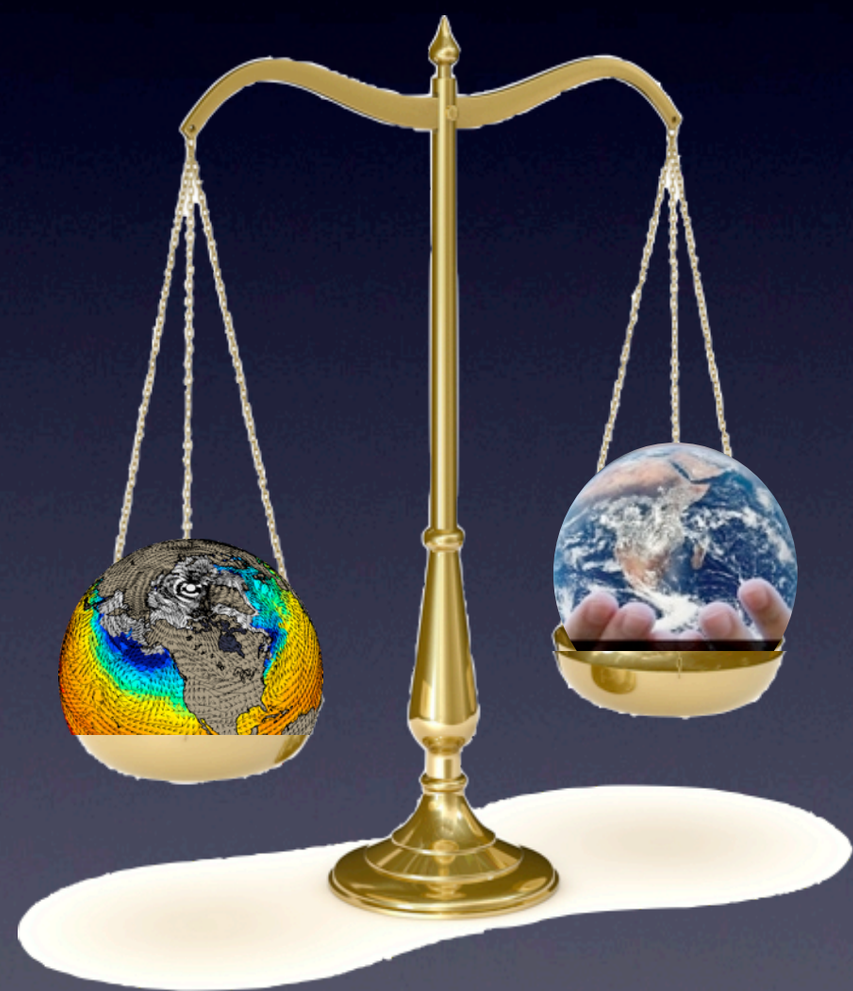
- **Community of Practice**
  Collaboratively develops data requirements, handling of scenarios

Ideal Climate Science Applications

80% Data handling

15% Analysis, Exploration, Uncertainty

5% Dissemination, Visualization



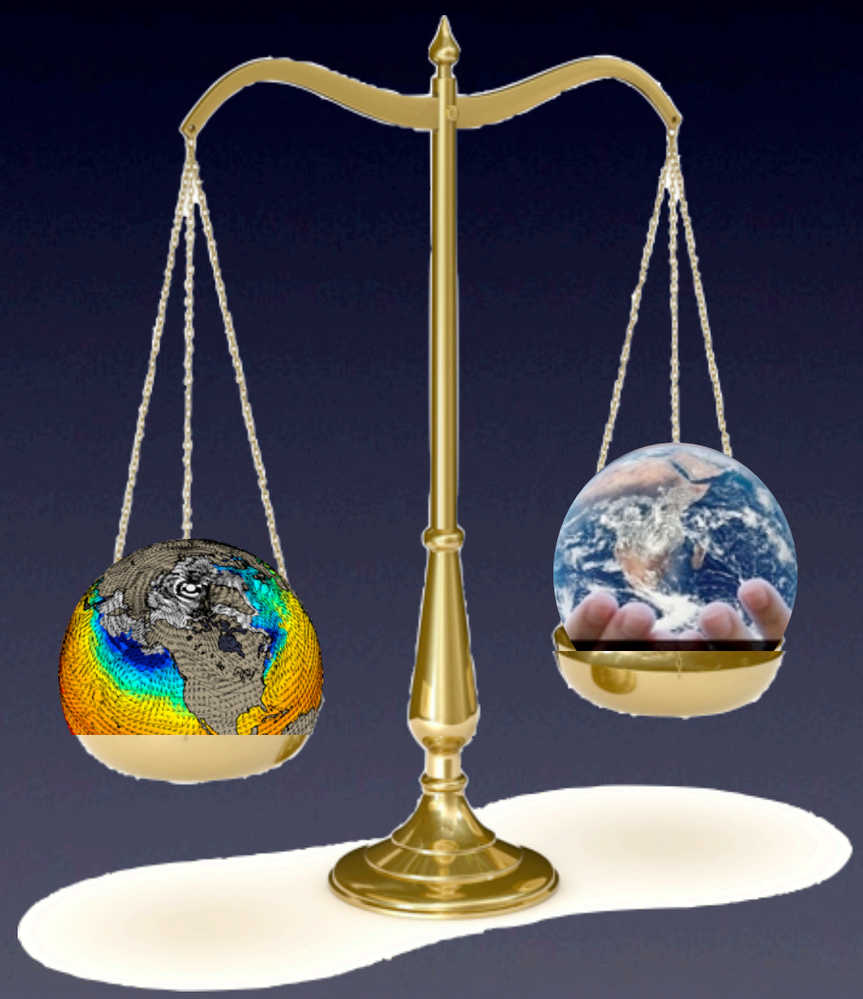
To be relevant, need to establish the right balance !

Ideal Climate Science Applications

~~80%~~ 10% Data handling

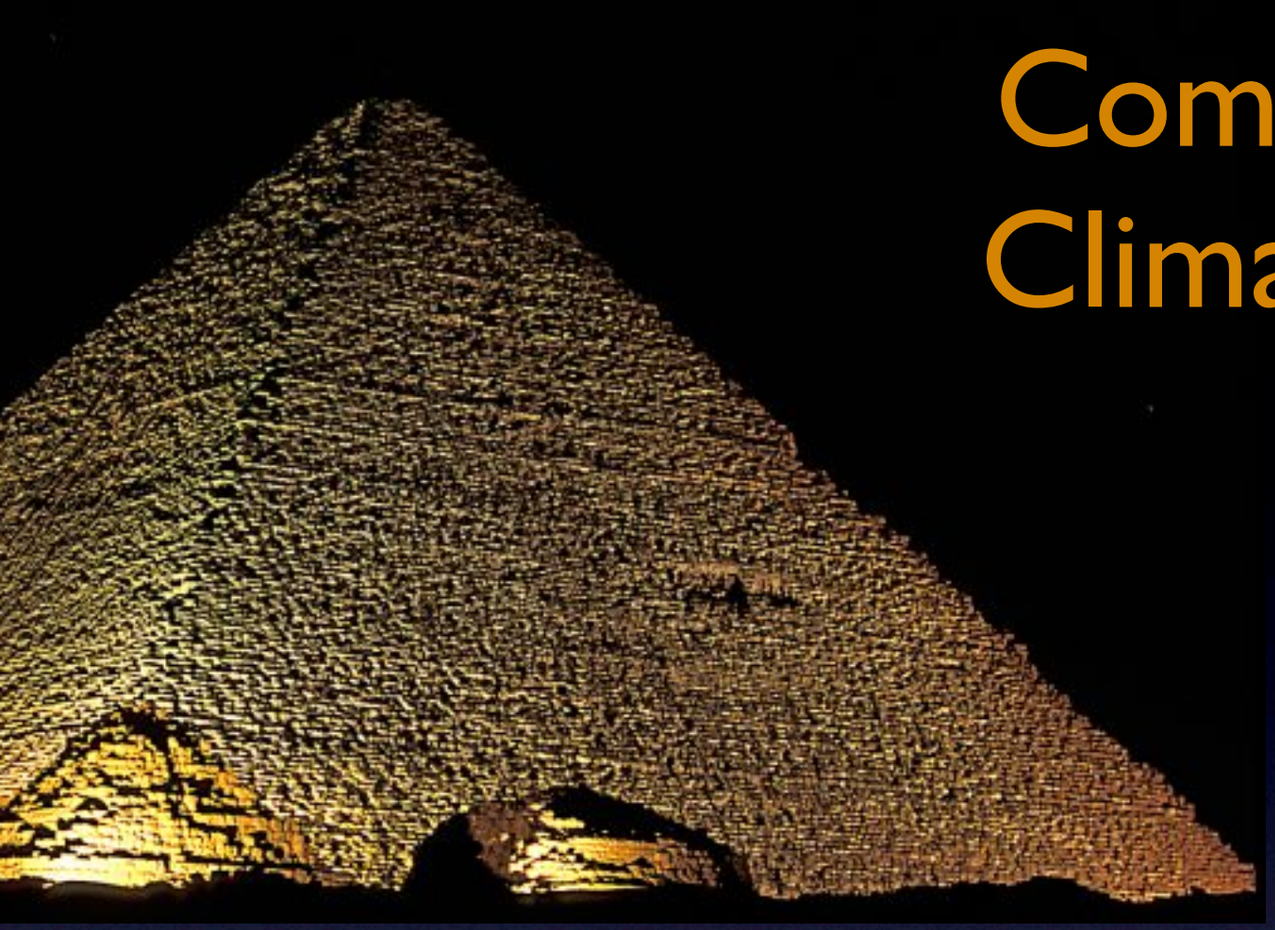
~~15%~~ 70% Analysis, Exploration, Uncertainty

~~5%~~ 20% Dissemination, Visualization



To be relevant, need to establish the right balance !

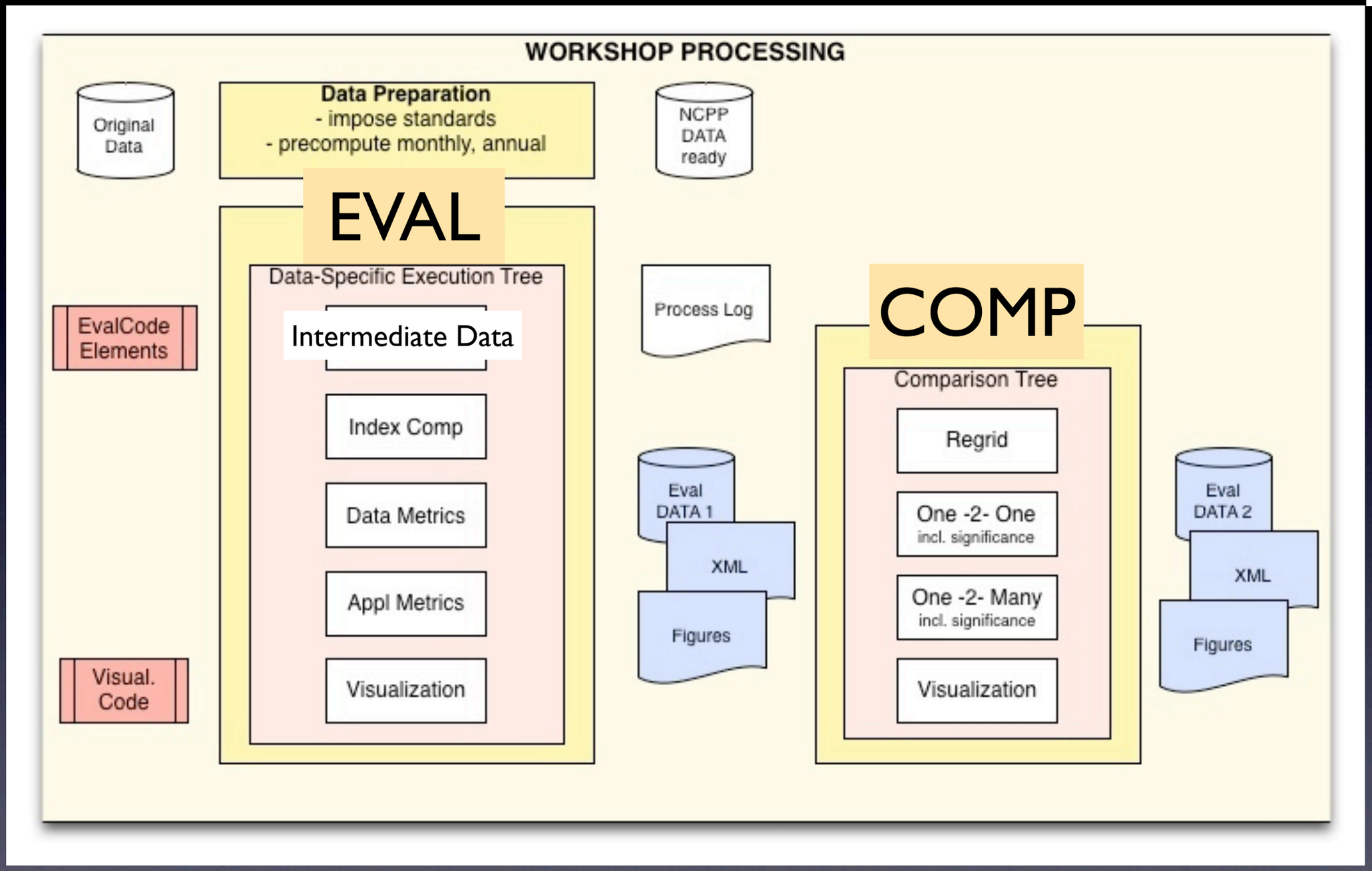
NCPP-Evaluation "Conveyor"



Components of a ClimateTranslator

- data access: based on services provided by UNIDATA, NOAA-NESII, DataCenters (NCDC, USGS, PCMDI, NCAR, U-Idaho, U-Del., U-Wisc., ...) : (workshop development: direct staging on NCAR Yellowstone)
- standardized indices for applications (CLIMDEX, BIOCLIM, NCA,...)
- Evaluation I: standardized characterization (protocols and metrics)
- Evaluation II: standardized intercomaprison options
- Foundation: extensible capabilities, documented standards

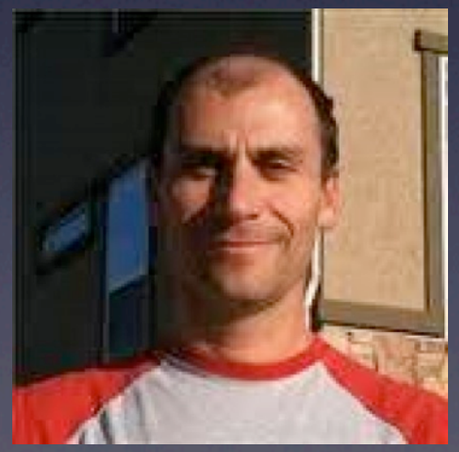
Plan of Evaluation Suite Processing



- Products: netCDF data, XML-Metadata, PNG-Visualization
- Open Source Software (ncl-repository), tested. *(availability via GitHub: planned)*
- Open access through EarthSystem CoG

Silent Stars of the QED v1.0

Jonathan Vigh
NCAR



Luca Cinquini
JPL / NOAA

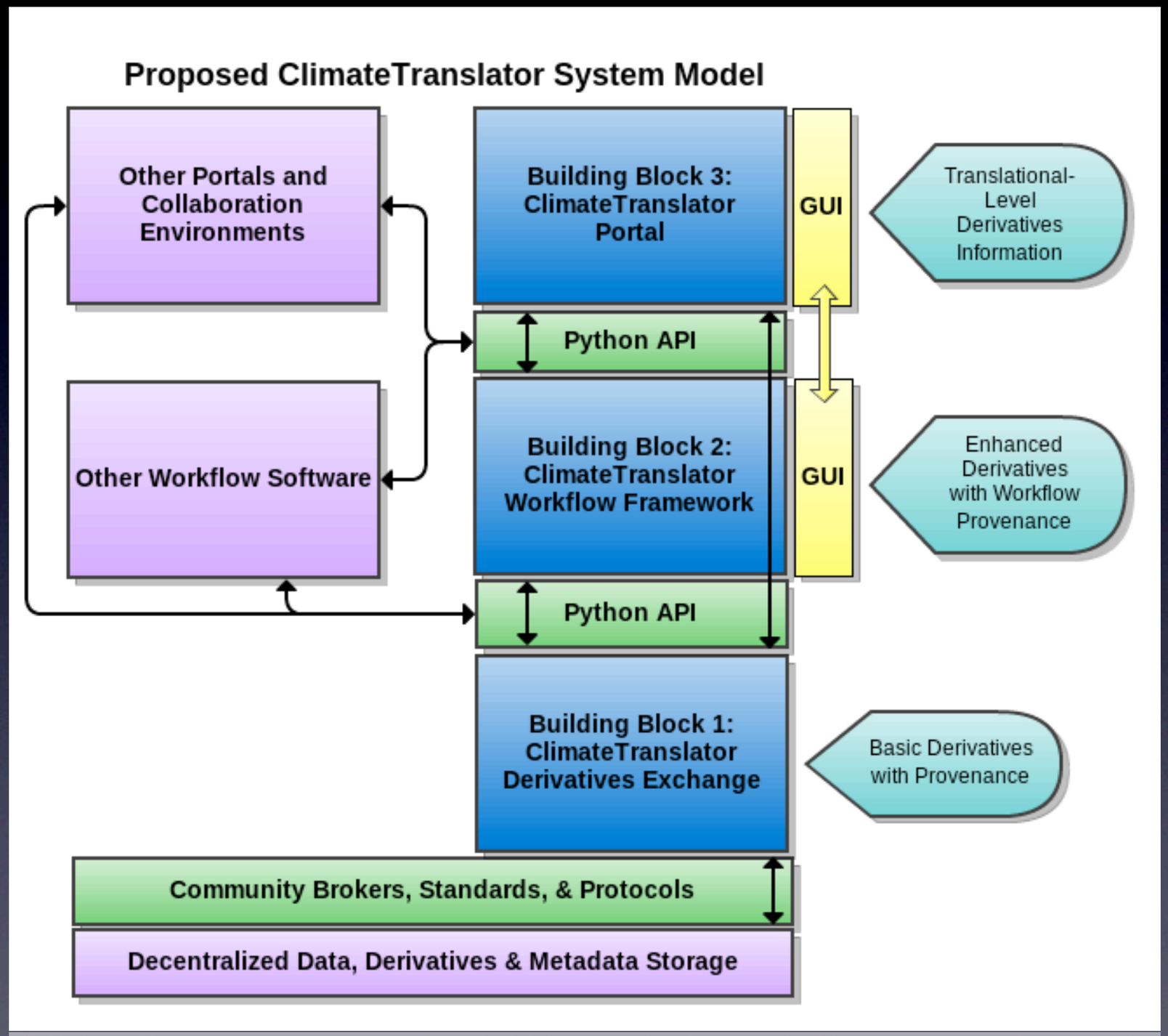
But also: DataCenters, Coord-Experiment Designers and their (silent) Stars!

ClimateTranslator by the Numbers

- 39 datasets (pr, tas, tasmin, tasmax, dtr)
- 17 variables and indices
- 80,000+ visualizations
- gridded observations, GCMs, dynamical and statistical downscaled data
- cleaned 30-yr present data_input: ~1TB
- Intermediate data: 2.5TB
- Output data (eval/comp): 0.3TB
- time from ingestion to plots: 2-10h

Data ID?

Full ClimateTranslator Vision



Translation Guidance

Evaluation Comparison

Data Service

Emerging Themes for Working Groups

1) "LinkedIn" for DATA / CODE: Topics in crowd-sourced data collection and workflow transparency: requires strict standards (NOAA-NCDC, DataONE, NCAR ...)

2) BETTER DATA: Observational data and downscaling tools to improve existing data products, and expand beyond precipitation and temperature variables. Stations, gridded data, bias correction? Multi-variate coherent data products? Multi-variate coherent data products?

3) DATA SERVICES: Cloud-based solutions for accessing, sub-setting, processing, and downloading resulting climate data products. Handling of large observational and model output data without having to deal with size, formats, ...

4) INSIGHTFUL COMPARISON: Tools for comparing high resolution climate model output with observational data for assessing applicability for impact studies

5) FACILITATING (study of) UNCERTAINTY: Approaches for expressing and applying uncertainty in gridded data products for use in subsequent impact studies. Sampling, processing uncertainty, and tools to handle output from ensembles and scenario generators for multi-model outputs

6) TUTORIAL: How to make sophisticated modeling tools, e.g., Bayesian Hierarchical Models accessible to application studies

7) RESOLUTION: Assess and document trades-offs between resolution and ensemble size for impact studies

Workshop Evaluation Objectives

- how to improve data access? which level of processed data?
- what additional variables, indices, metrics? Prioritization?
  but what is an observational reference?
- what pre-computed visualizations are useful? when only data?
- is this infrastructure useful to build guidance?
  what can be added/changed to support your communities?