International Conference on
Atmospheric Chemical Mechanisms
UC Davis, December 7, 2006

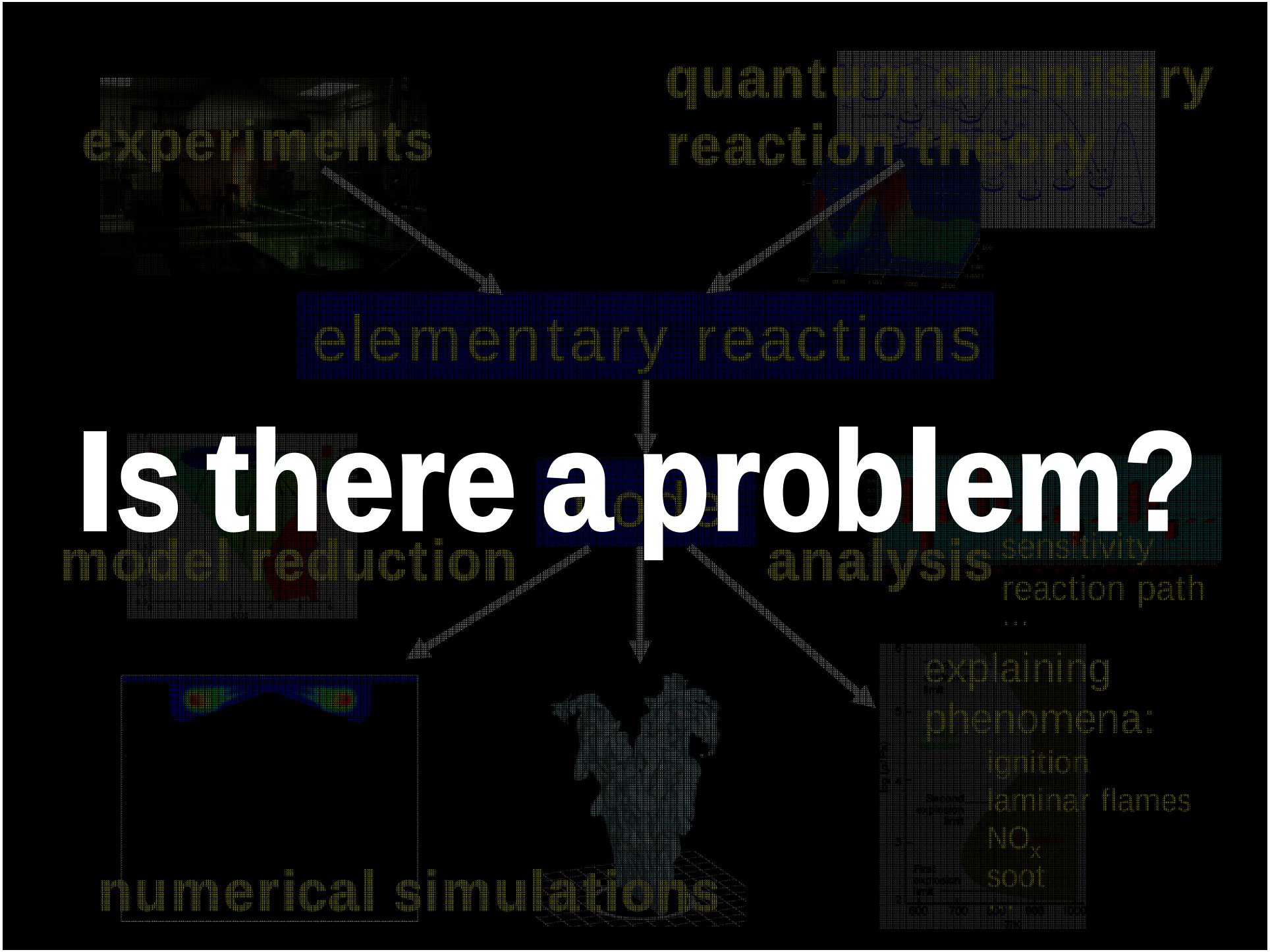
Transforming Data into Knowledge:
Process Informatics for Combustion Chemistry

M. Frenklach, A. Packard, Z.M. Djurisic, R. Feeley, T. Russi (UCB)
D.M. Golden, C.T. Bowman (Stanford)
W.H. Green, Jr., G.J. McRae (MIT)
P.J. Smith, A. Sarofim (U of Utah)
H. Wang (USC)
G. Smith (SRI)
P.R. Westmoreland (U of Mass)
T.C. Allison (NIST)
J. G. Frey, M.J. Pilling (Leeds)
. . .

Supported by:
NSF Chemistry Division (Cyber-infrastructure)

Dmitri Mendeleyev

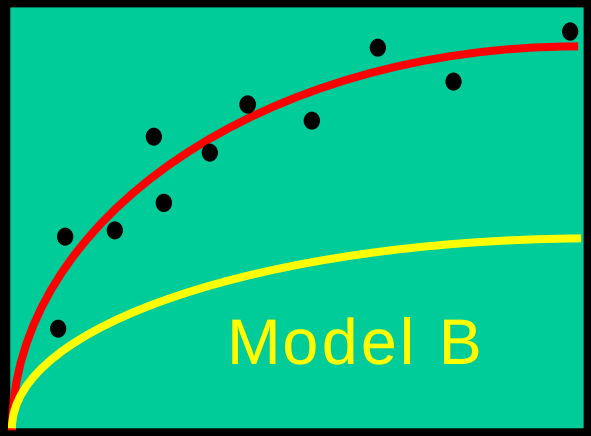
"The mere accumulation of facts, even an extremely extensive collection, ... does not constitute scientific method; it provides neither a direction for further discoveries nor does it even deserve the name of science in the higher sense of that word. The cathedral of science requires not only material, but a design, harmony ... a design ... for the harmonic composition of parts and to indicate the pathway, by which the most fruitful new material might be generated."



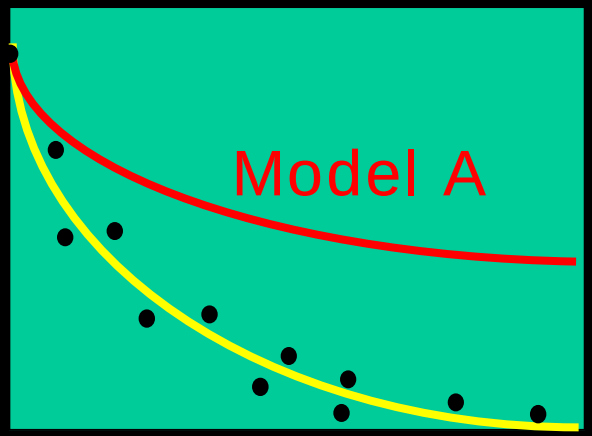
Is there a problem?

A TYPICAL SEQUENCE OF EVENTS

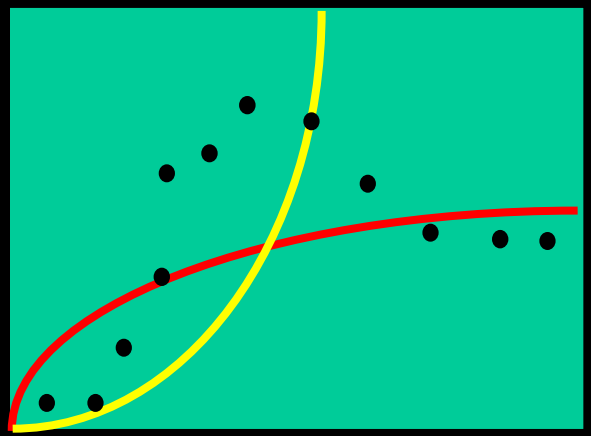
Experimenter A



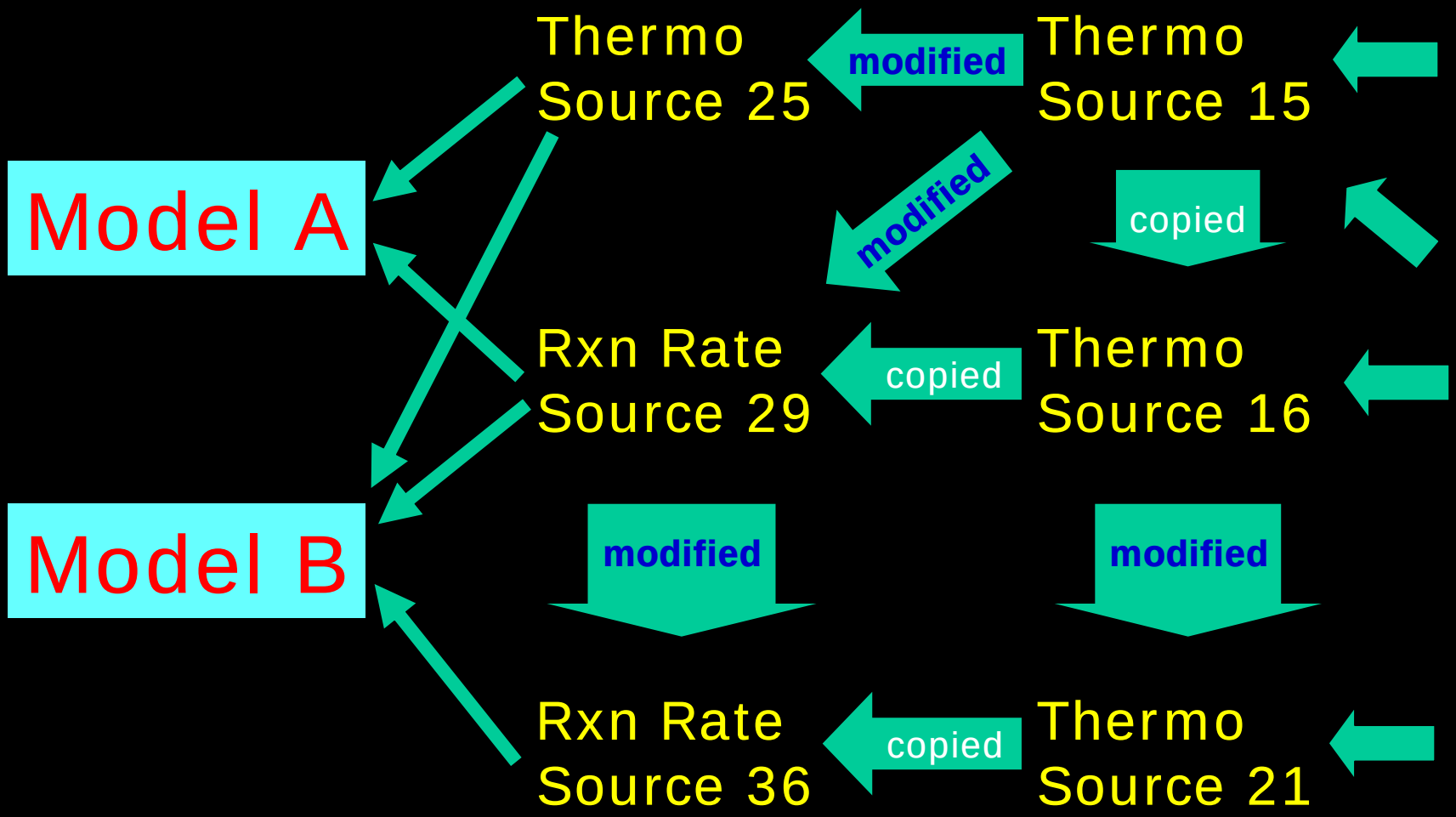
Experimenter B



Experimenter C



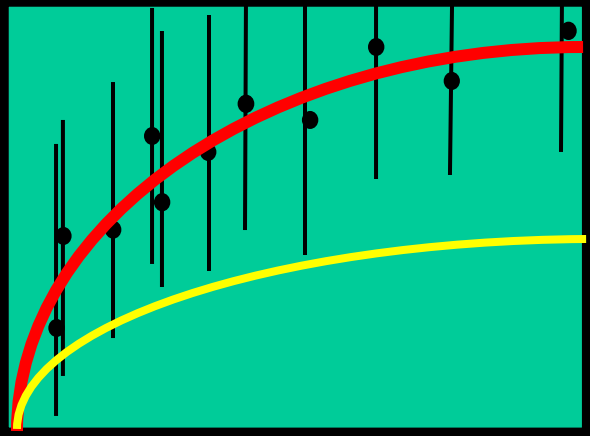
AND CHAOS OF DATA



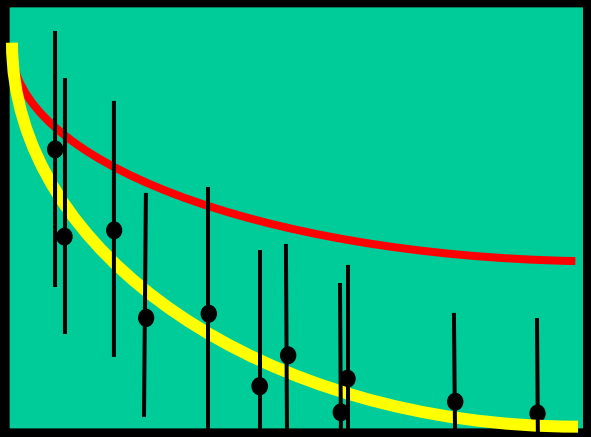
PROBLEMS

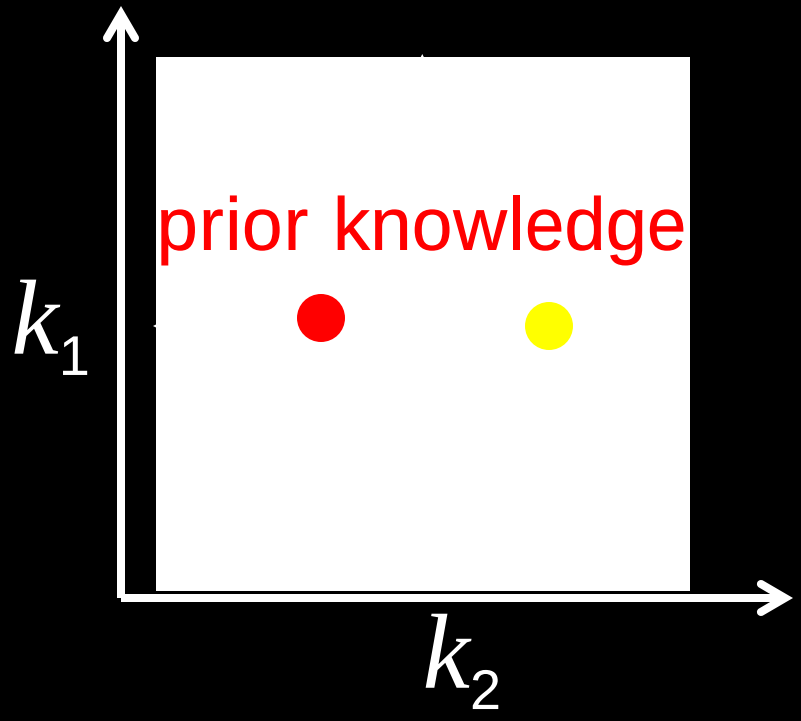
- multitude of sources
- conflicting data in/among sources
- poor documentation of data
- no uncertainty reporting or analysis
- not much focus on integration of data

Experimenter A

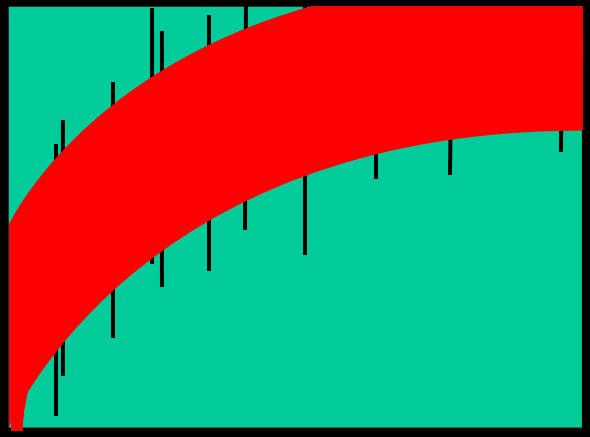


Experimenter B

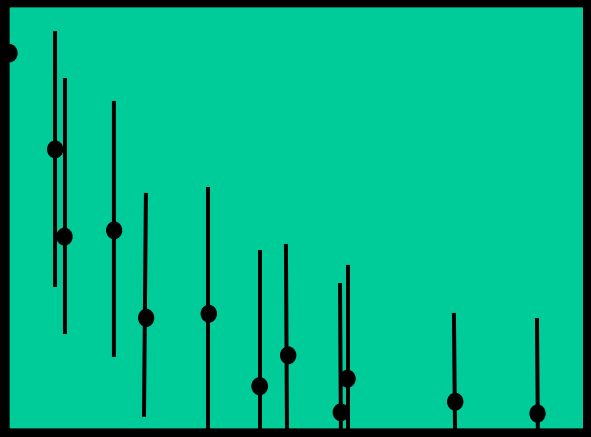


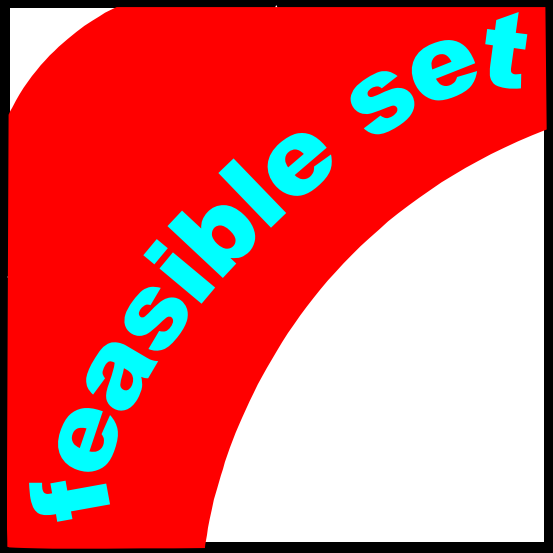


Experimenter A

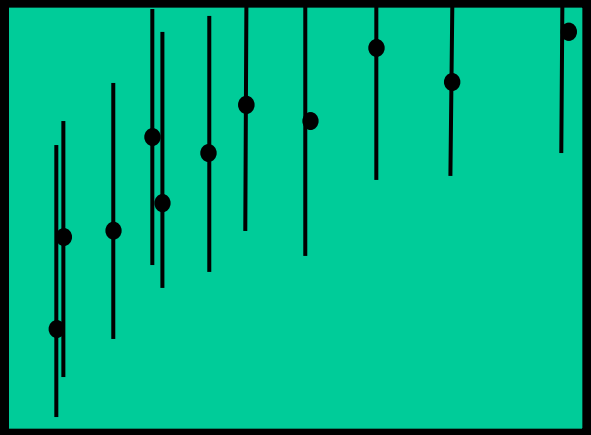


Experimenter B

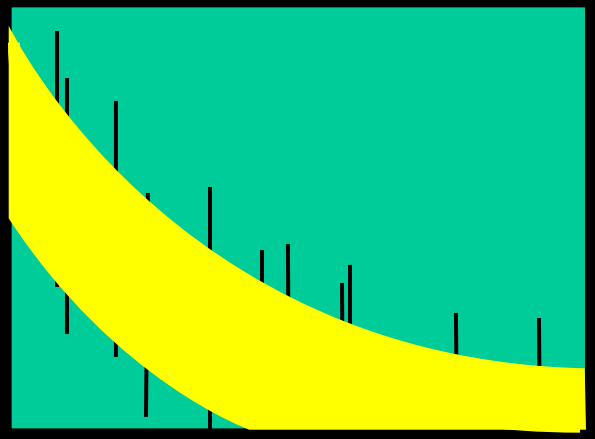


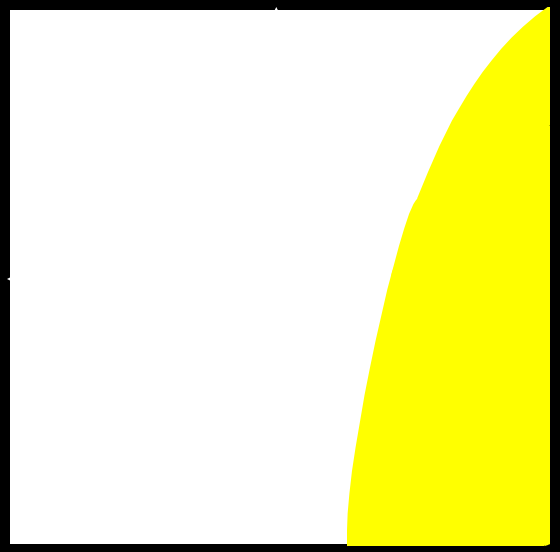


Experimenter A

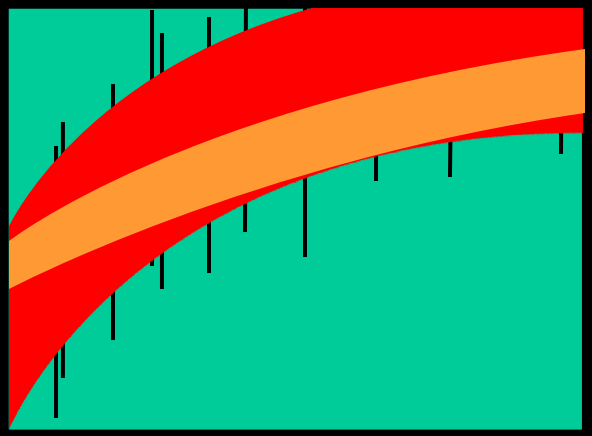


Experimenter B

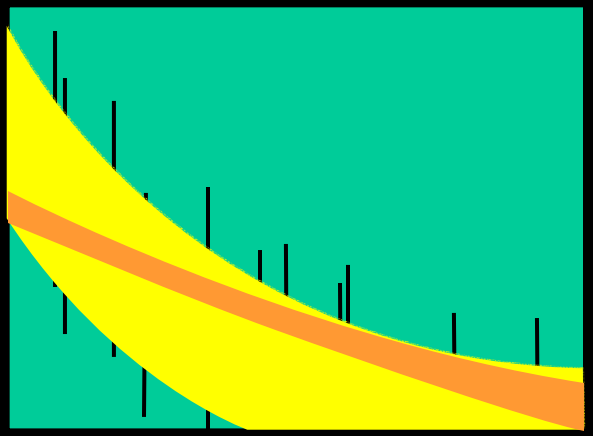


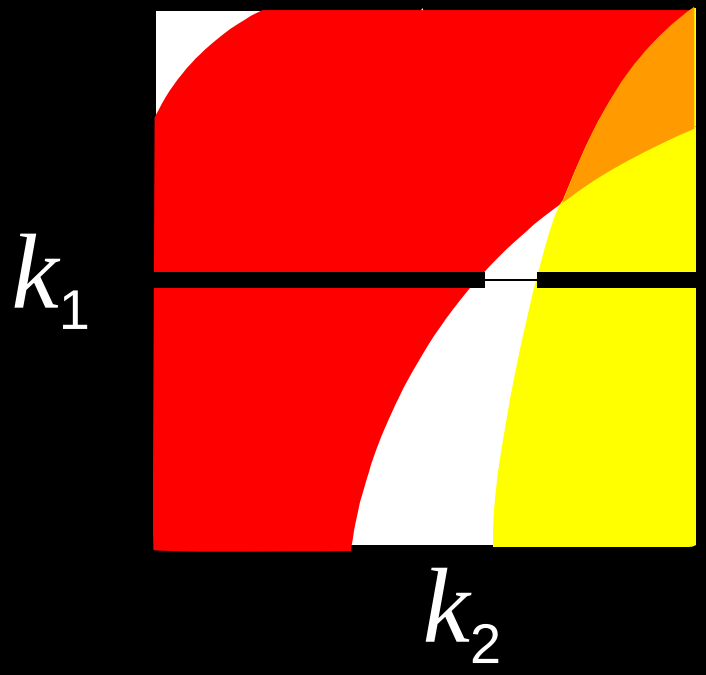


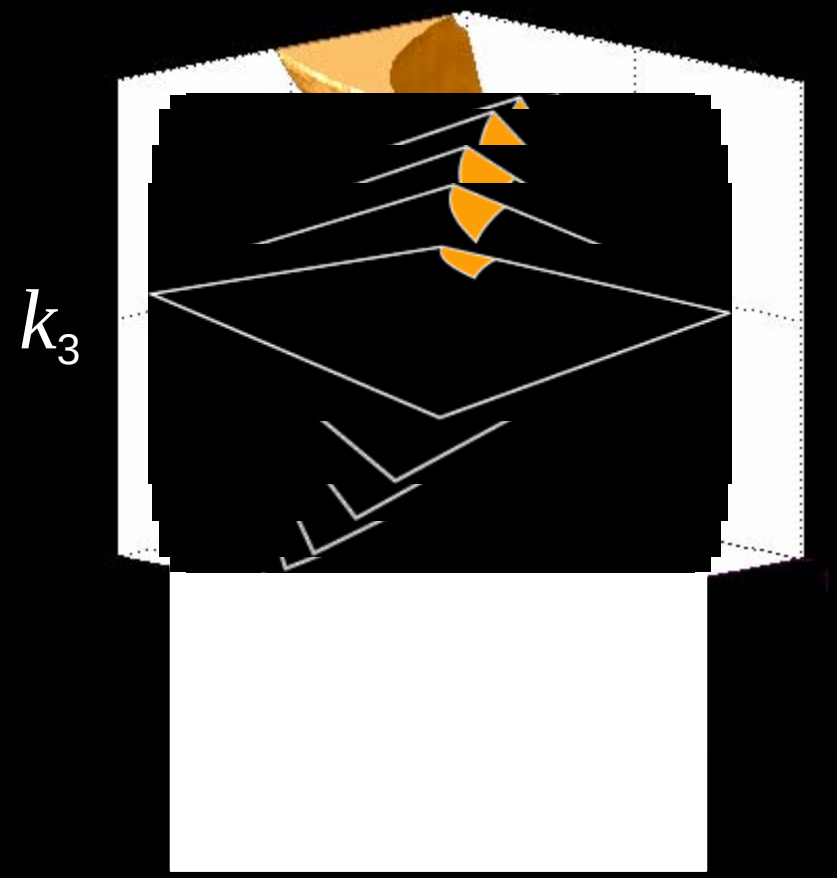
Experimenter A



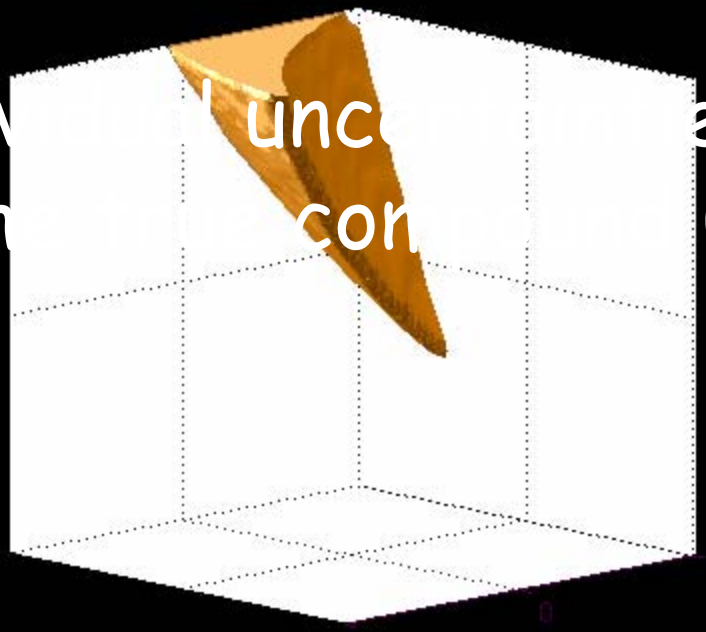
Experimenter B






k_3

- an "expert choice" of parameter values is like a needle in a haystack
- a set of indiv[illegible] unc[illegible]es does not represent th[illegible] con[illegible] uncertainty



- models are not additive
- data are not additive
- need a *system*
  for synthesis and transformation
  of data into knowledge

PROCESS INFORMATICS SYSTEM ?

- Mechanism understanding
  through experimentation and modeling
  for the purpose of prediction
- Integrated system of data and tools – System Science
- Collaborative science
- Applications to:
  - chemical kinetics
  - combustion
  - atmospheric chemistry
  - astrophysics
  - system biology
  - …

PrIMe

PrIMekinetics.org

Process Informatics Model

THEME

The community builds PrIMe,

PrIMe builds the community

Sometime in 2008 …

Scientist to SYSTEM:

What impact will my experiment have on…?

SYSTEM to Scientist:

Only if you control error within 2 % …

Sometime in 2010 …

Engineer to SYSTEM:

I need a reliable prediction for conditions…

SYSTEM to Engineer:

The best one will give error bars of 30 % …
$$ can drop it to 10 %, and $$$$ to 3 %

Sometime in 2020 ...

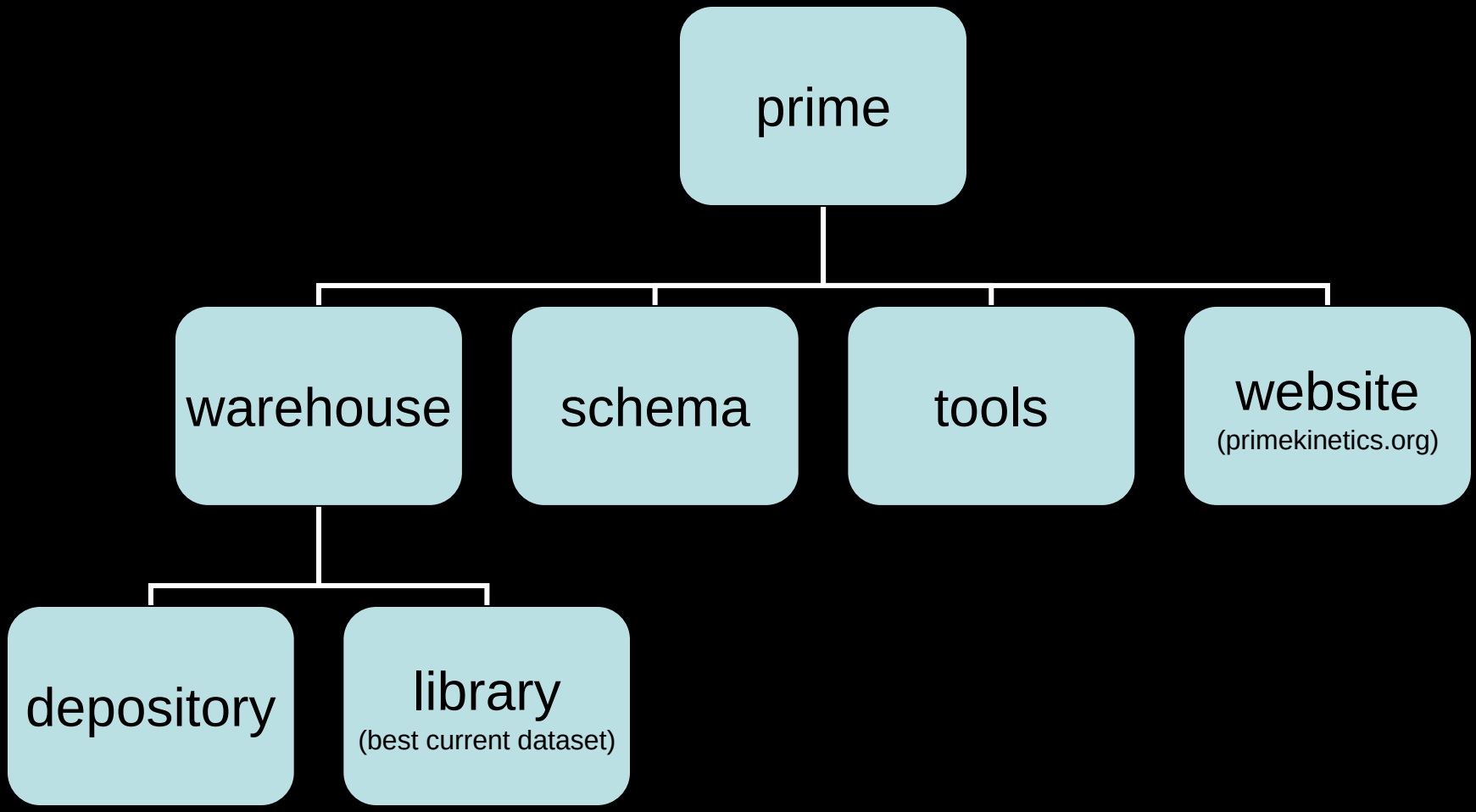
Policymaker to SYSTEM:

How much longer will there be an Antarctic ozone hole?

SYSTEM to Policymaker:

. . .

PrIMe INFRASTRUCTURE



PrIMe INFRASTRUCTURE

- Data management
- Web services
- Model generation
- Numerical solution (flames, flow reactors, …)
- Model reduction
- Optimization
- Error analysis/propagation
- . . .

tools

DATA COLLABORATION

– model/data sharing and automated reasoning to extract the totality of information in the community data

- Dataset: a set of {Observation, Uncertainty, Model}
- Prior Information on parameters
- Assertions/prediction over the feasible set
- Use global optimization methods, from control systems analysis, convex analysis, and algebraic geometry

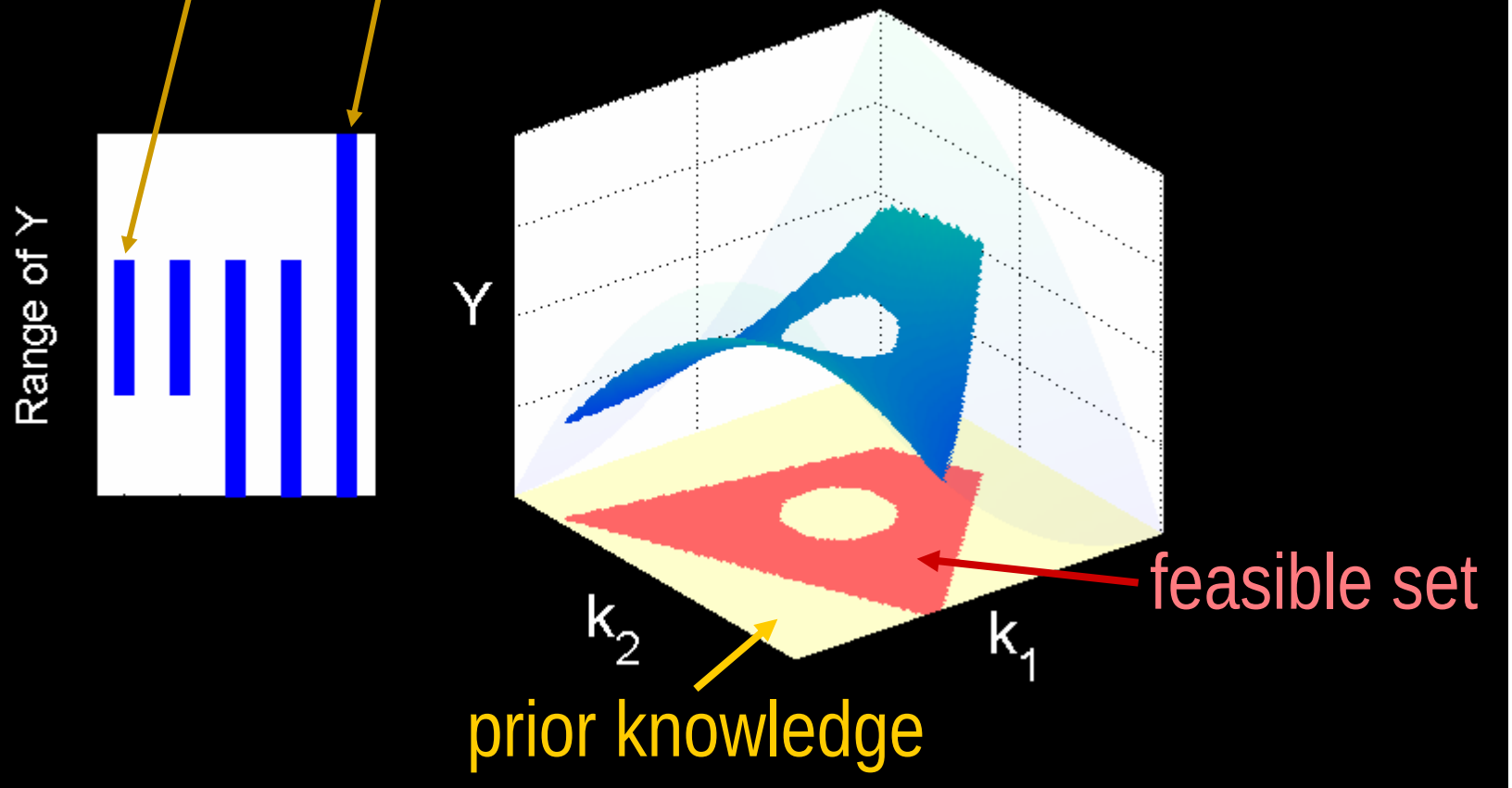
PREDICTION ON THE FEASIBLE SET

Information Gain:

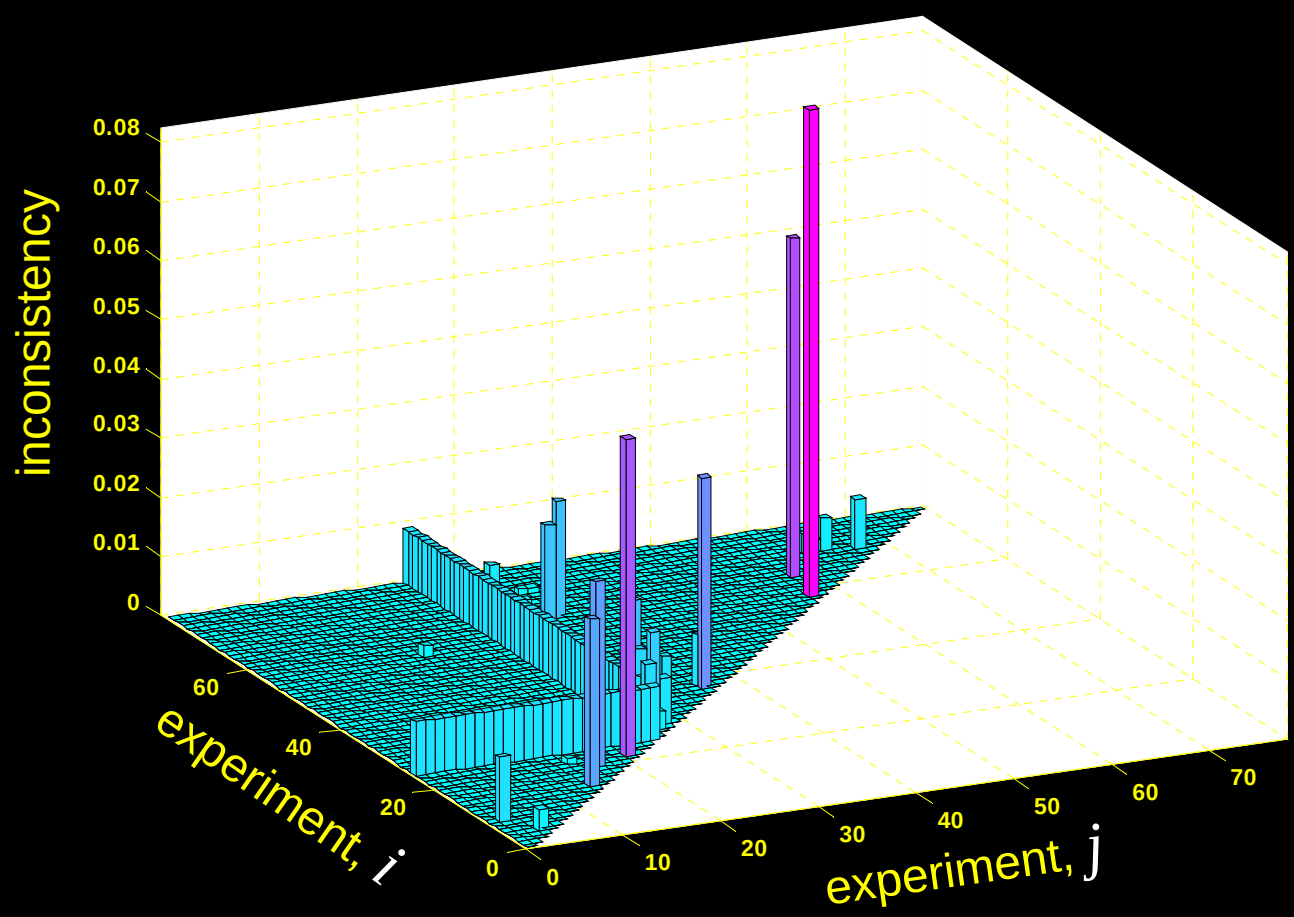
$$I = 1 - \frac{\text{Posterior Range}}{\text{Prior Range}}$$

Only Prior Knowledge

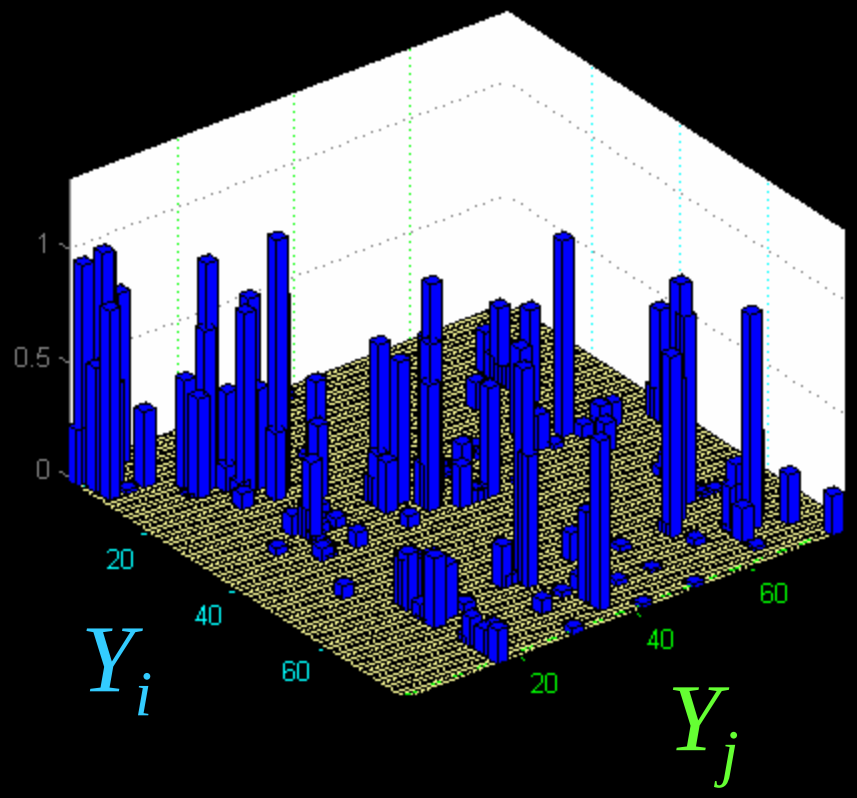
Including Experimental Observations



Pair-wise consistency test of GRI-Mech 3.0 Dataset



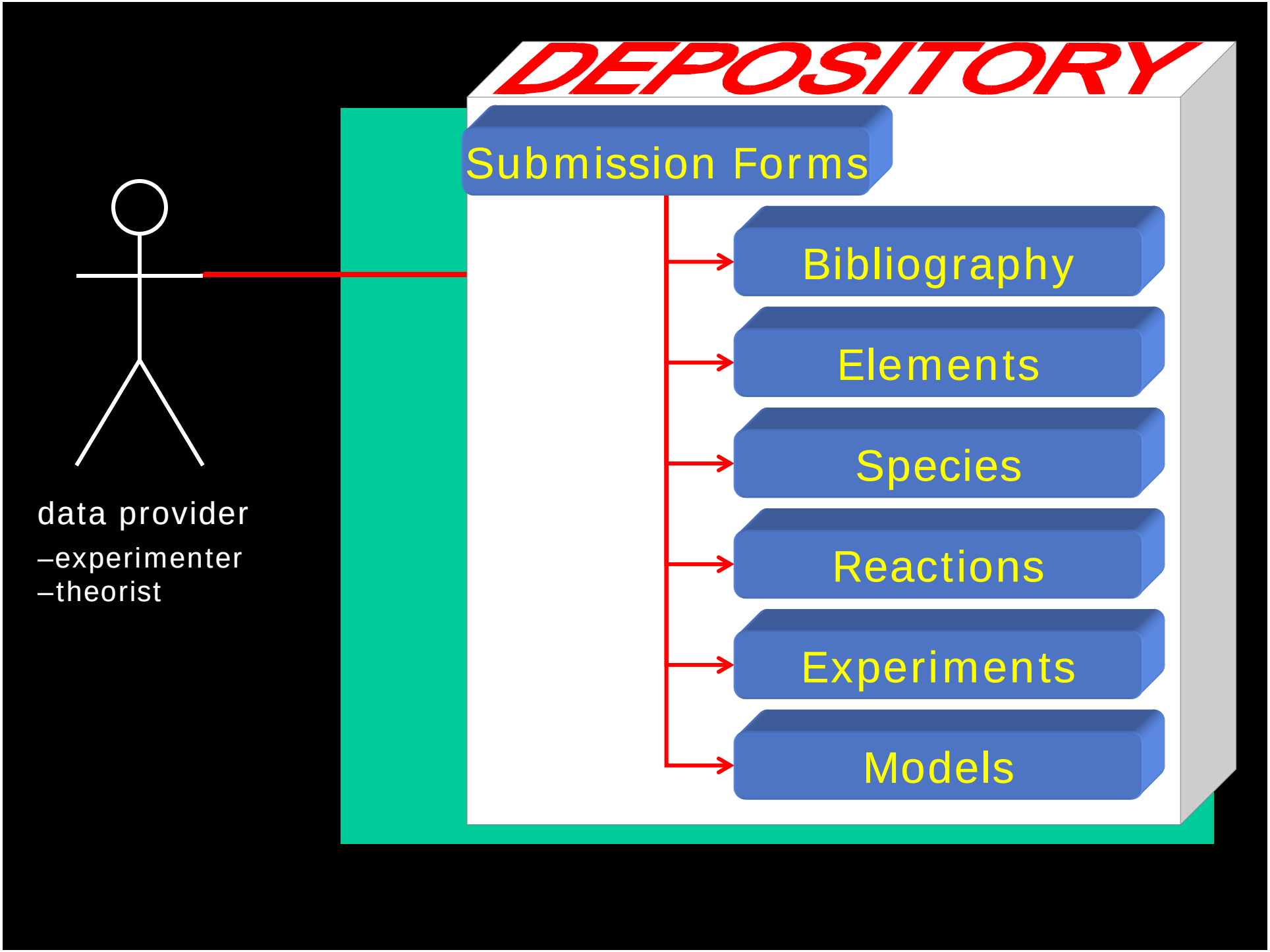
Sensitivity $\dfrac{\text{of uncertainty in predicting } Y}{\text{to uncertainty in observation } Y_j}$

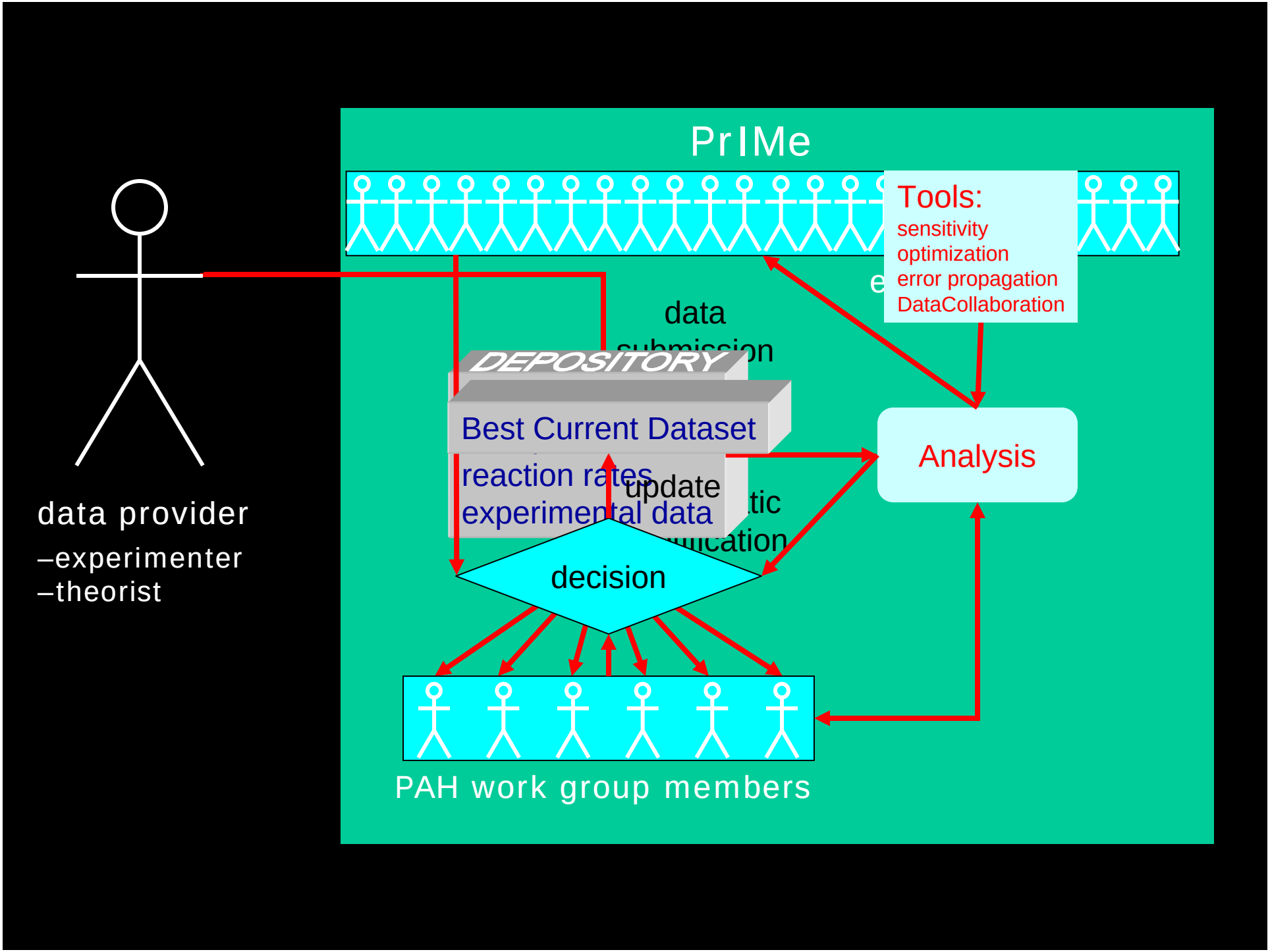


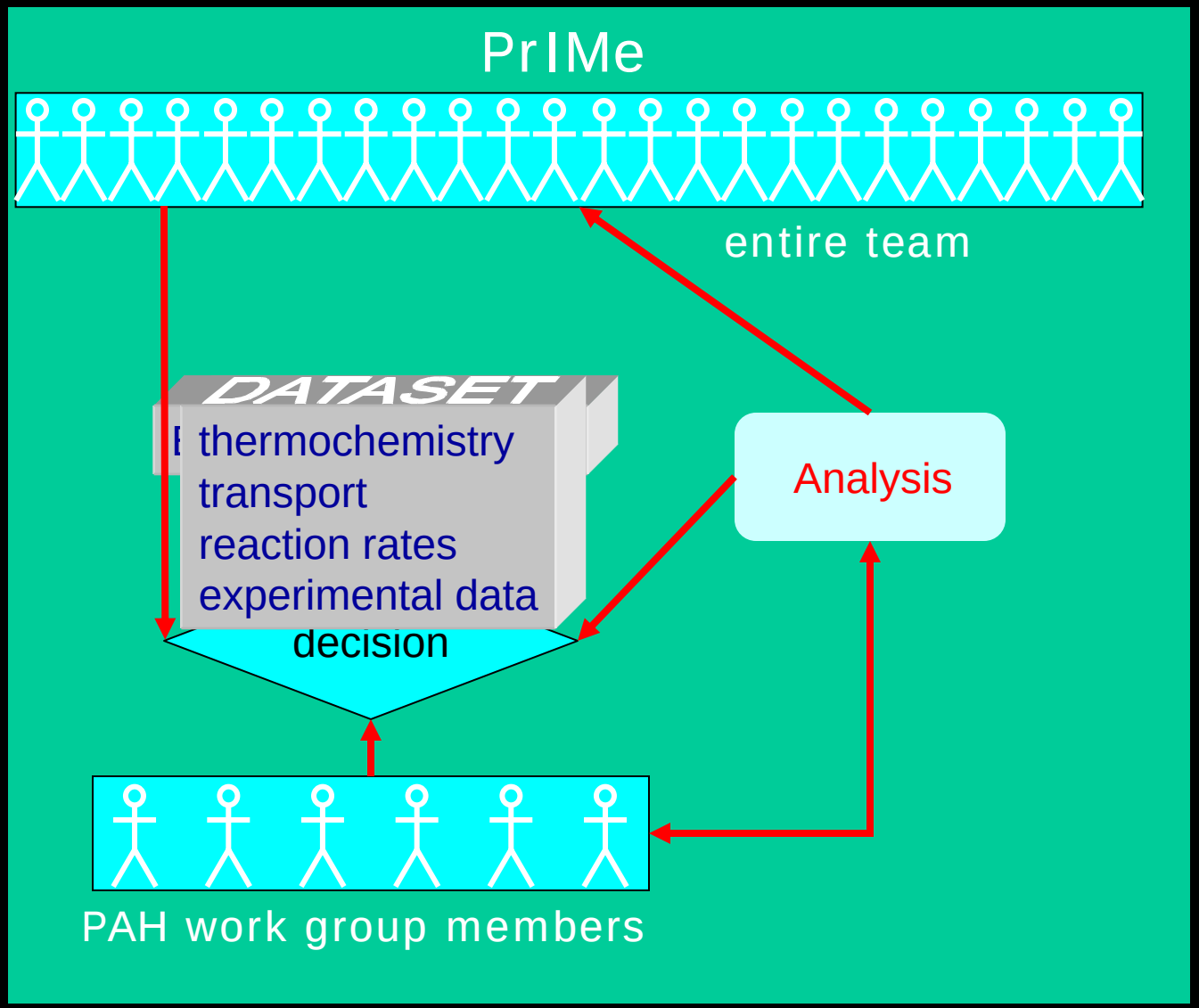
PrIMe organization:

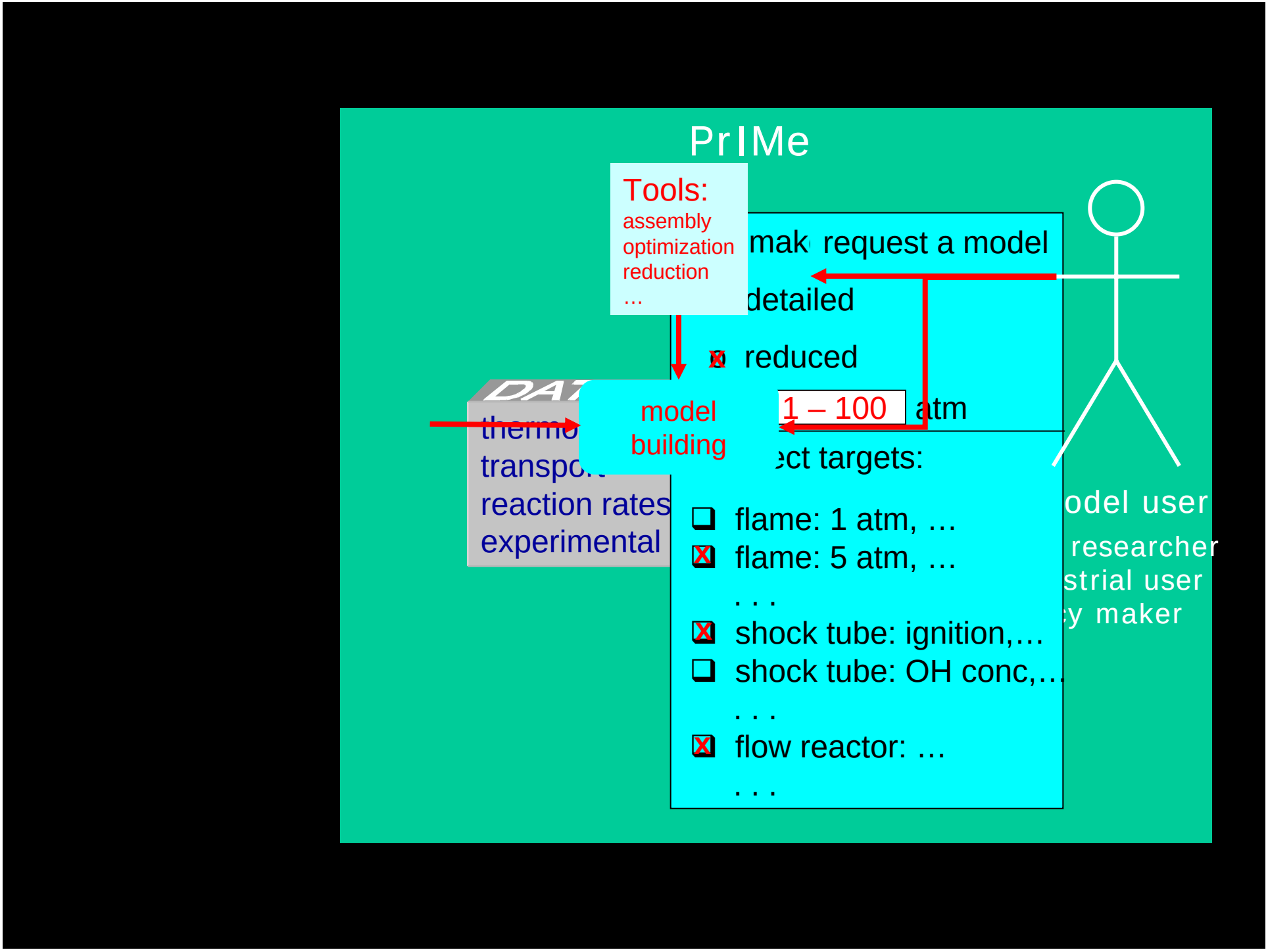
practical realization of collaborative data management

```
<?xml version="1.0" encoding="UTF-8" ?>
- <experiment primeID="x00000037" xmlns="http://purl.org/NET/prime/"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://purl.org/NET/prime/
    http://purl.org/NET/prime/schema/experiment.xsd
    http://purl.org/dc/elements/1.1/
    http://dublincore.org/schemas/xmls/qdc/2003/04/02/dc.xsd">
    <copyright>©primekinetics.org 2005</copyright>
    <bibliography preferredKey="Luque et al., 1996"
      type="primeBiblioID">b00014617</bibliography>
  - <apparatus>
      <kind>flame</kind>
      <mode>laminar</mode>
      <mode>premixed</mode>
    </apparatus>
  - <commonProperties>
    - <property name="pressure" units="atm" label="P" description="flame
        pressure">
        <value>1</value>
      </property>
    + <property name="flow rate" units="g/(cm² s)" label="m_dot"
        description="mass flow rate">
    + <property name="initial composition">
    </commonProperties>
  - <dataGroup id="dg1" label="CH profile">
      <property id="x1" name="distance" units="cm" label="x" description="height
        above burner" />
    - <property id="x2" name="concentration" units="molecules/cm³" label="[CH]"
        description="CH concentration">
        <species preferredKey="CH" type="primeSpeciesID">s00009010</species>
      </property>
    - <dataPoint>
        <x1>0.077154</x1>
        <x2>2.08e+010</x2>
      </dataPoint>
    - <dataPoint>
        <x1>0.080385</x1>
        <x2>2.16e+010</x2>
      </dataPoint>
    - <dataPoint>
        <x1>0.083615</x1>
        <x2>2.3e+010</x2>
      </dataPoint>
    - <dataPoint>
        <x1>0.086846</x1>
```


DEPOSITORY
Submission Forms
Bibliography
Elements
Species
Reactions
Experiments
Models
data provider
–experimenter
–theorist


PrIMe
Tools:
sensitivity
optimization
error propagation
DataCollaboration
data
DEPOSITORY
Best Current Dataset
reaction rates
experimental data
update
Analysis
decision
data provider
–experimenter
–theorist
PAH work group members

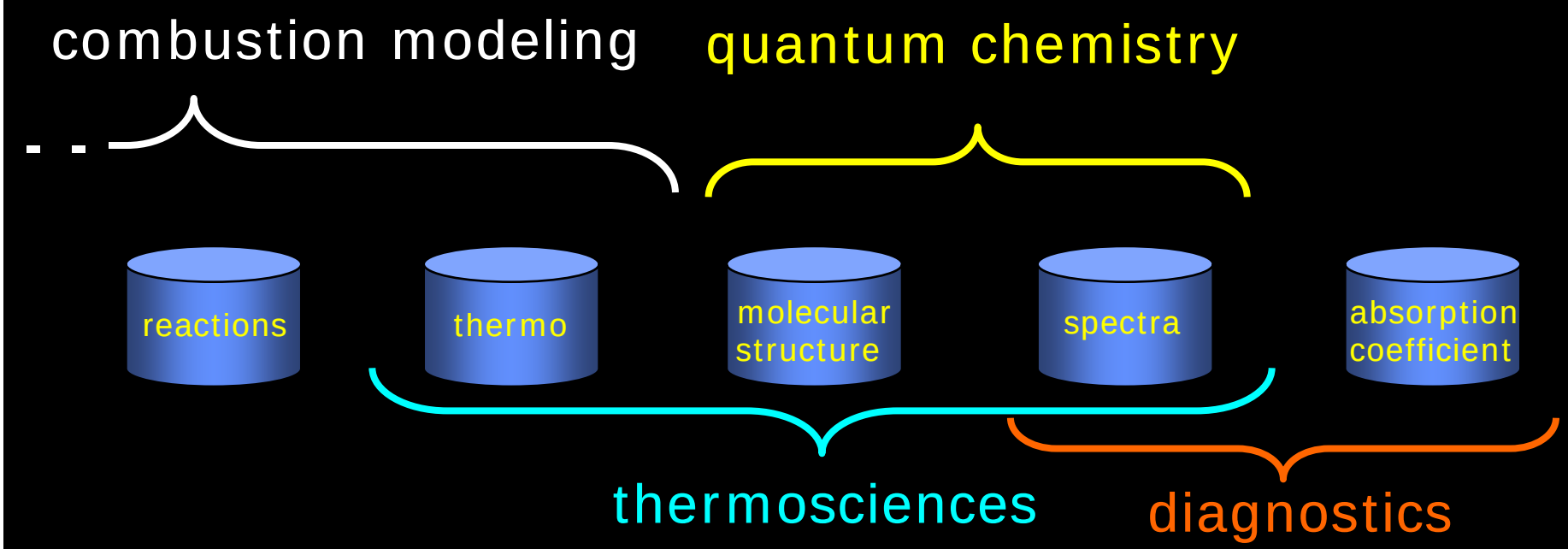

PrIMe
entire team
DATASET
thermochemistry
transport
reaction rates
experimental data
decision
Analysis
PAH work group members


PrIMe
Tools:
assembly
optimization
reduction
...
mak request a model
detailed
reduced
model
building
1 – 100 atm
ect targets:
flame: 1 atm, …
flame: 5 atm, …
. . .
shock tube: ignition,…
shock tube: OH conc,…
. . .
flow reactor: …
. . .
reaction rates
experimental
odel user
researcher
strial user
y maker

BENEFITS of PrIMe

- establishes/advances system science
  - synthesis of disparate data into shared knowledge
  - crosses and links all physical scales
  - expandable and extendable, modular and parallelizable
  - identifies and ranks problem areas and tests possible action
- addresses group sociology
  - opportunity and incentive for scientific collaboration within and among communities (“virtual” collaboration)
  - means for reaching broad community consensus
  - inclusive, yet “filters out” bad science
  - respects and protects individual territory
- enables technology transfer
  - continuous/instantaneous, in parallel with science
  - solid base for technology development
  - responsiveness to (individual) technological needs
  - assures privacy if so desired
- leverages resources
  - distributed funding — funding of individual Work Groups

MODULARITY, EXTENDIBILITY, ...



CURRENT STATUS OF PrIMe

- grass-roots initiative
- open membership
- launched on April 21-22, 2006
- data Depository
  - NIST kinetic database
  - GRI-Mech 3.0
  - Stanford “Volume 1”
  - USC flames, EU/Leeds kinetic database, …
- 20 Work Groups – all working in parallel
  - reaction subsets, thermodynamics, transport
  - experimental data
  - cyber-numerics (Web services, Grid)
  - industrial
- 88 registered members (as of December 6)
  from 14 counties

CURRENT STATUS OF PrIMe

- data was moved to “24/7” service
- initial release of PrIMe Data Management software is within days
- next infrastructure milestone: Web services
  - Grid access (UofUtah, Argonne)
  - Data Collaboration (UCB)
  - automated model generation (MIT)
  - Active Thermodynamic Tables (Argonne)
  - Sensitivity analysis (Hungary)
- first science “release”: C0-2 reaction system
  - communal collection and (re)evaluation of pertinent data
  - fully-automated, on-the-fly data analysis and model building
- IJCK/Wiley agreed to have a “direct” link to PrIMe
  - submission of data is tied to submission of manuscripts
  - automated assessment of newly submitted data, reporting to the author and reviewers