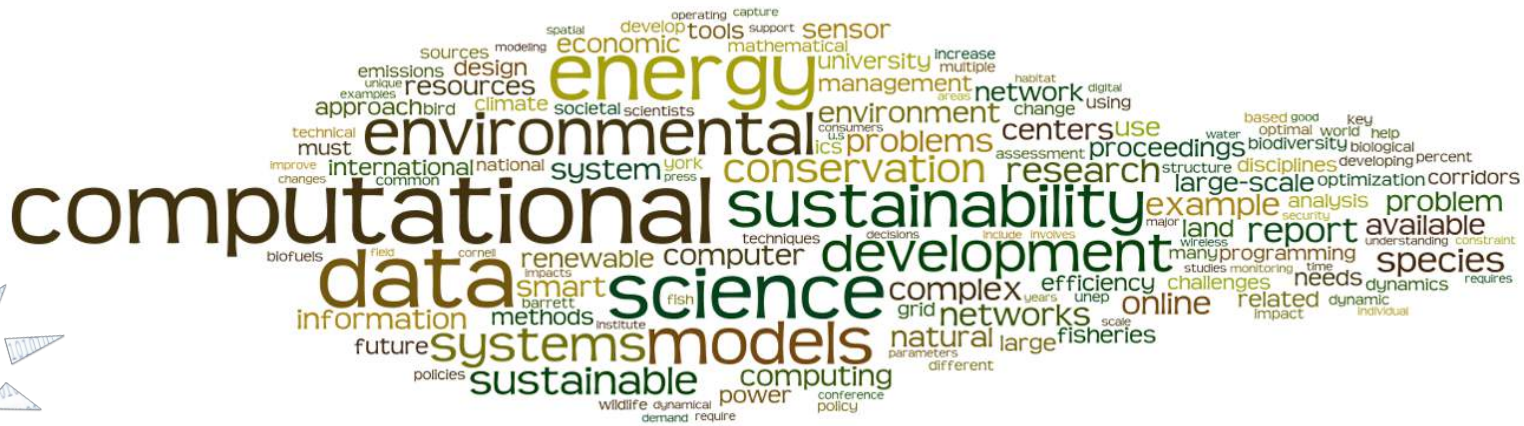
Computational Sustainability:
Computational Methods for Sustainable Development

Carla P. Gomes
Cornell University



Computing Research:
Addressing National Priorities and Societal Needs

CCC

May 2016

Expeditions in Computing (CISE)

Thank you!

Expeditions in Computing (CISE)

1st Expedition in Computational Sustainability (2008)

Conference, Referred CompSust Tracks, Workshops

- **To nucleate** the Computational Sustainability field
- **To identify** a number of **core research directions** for **maximal impact**, both in terms of **Computer Science** and **Sustainability.**

Sustainability is an ***uncharted territory for computer scientists,*** highly interdisciplinary area.

To ensure that the field of computational sustainability becomes self-sustaining

⇒

Critical mass and visibility that cannot be achieved with piece-meal research efforts.

2nd Expedition (2016)

CompSustNet: Large-Scale Research Network

Expanding the Horizons of Computational Sustainability

Computational Sustainability

New interdisciplinary field that aims to advance computational methods for Sustainable Development.

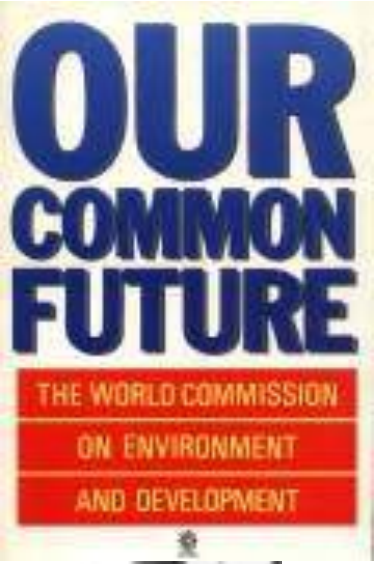
Sustainability and Sustainable Development

The 1987 UN report, "Our Common Future" (Brundtland Report):

- Raised serious concerns about the State of the Planet.
- Introduced the notion of sustainability and sustainable development:

> *Sustainability: "development that meets the needs of the present without compromising the ability of future generations to meet their needs."*

- Stated the urgency of policies for sustainable development.



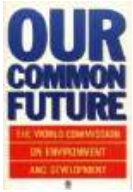
Gro Brundtland
Norwegian Prime Minister
Chair of WCED

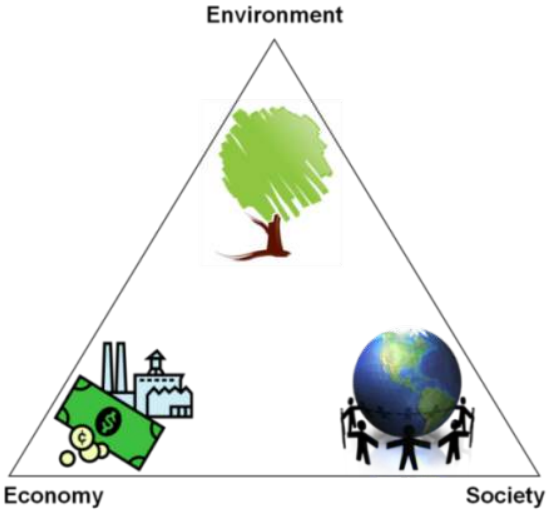
UN World Commission on Environment and Development,1987.

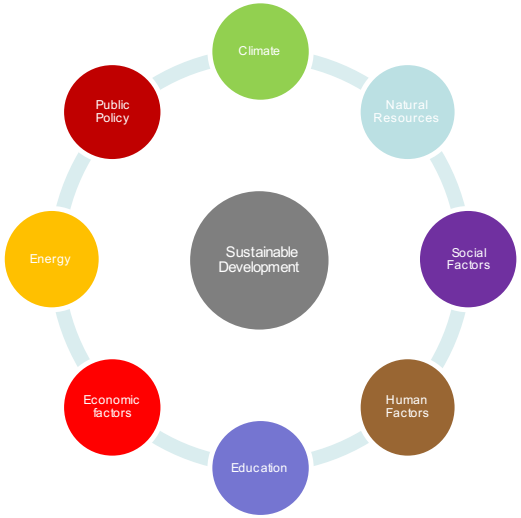
Sustainability: Interlinked environment, economic, and social

Sustainable Development encompasses balancing

- *environmental,*
- *economic, and*
- *societal needs.*





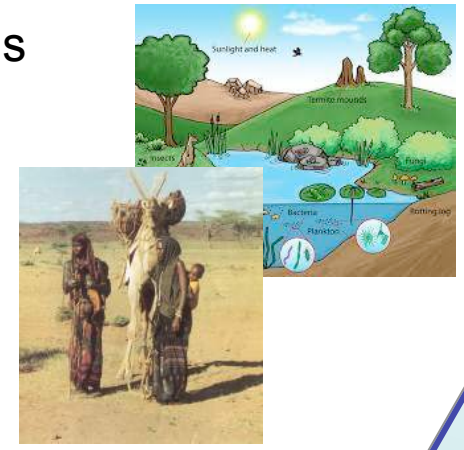


Ultimate goal of Sustainable Development
HUMAN WELL-BEING
of current and future generations.

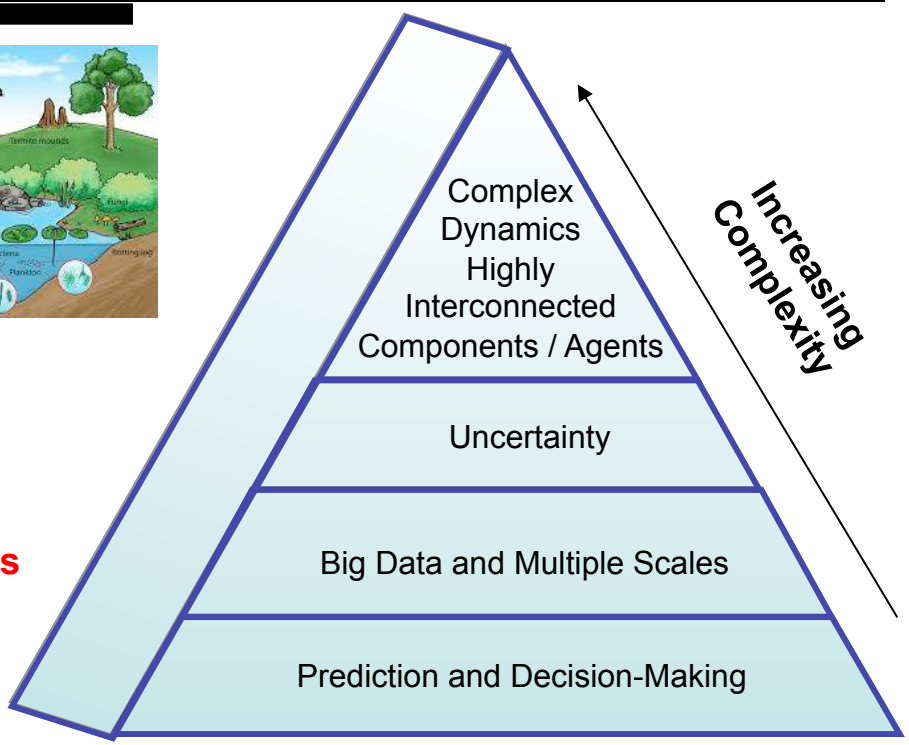
Sustainability Problems: Complex Systems

Sustainability problems
unique in scale
and complexity

Smart Power Grid:
Complex Digital Ecosystem



Natural Ecosystems



Significant Computational Challenges

Key challenge:
How to effectively establish the
necessary large-scale interdisciplinary projects and collaborations

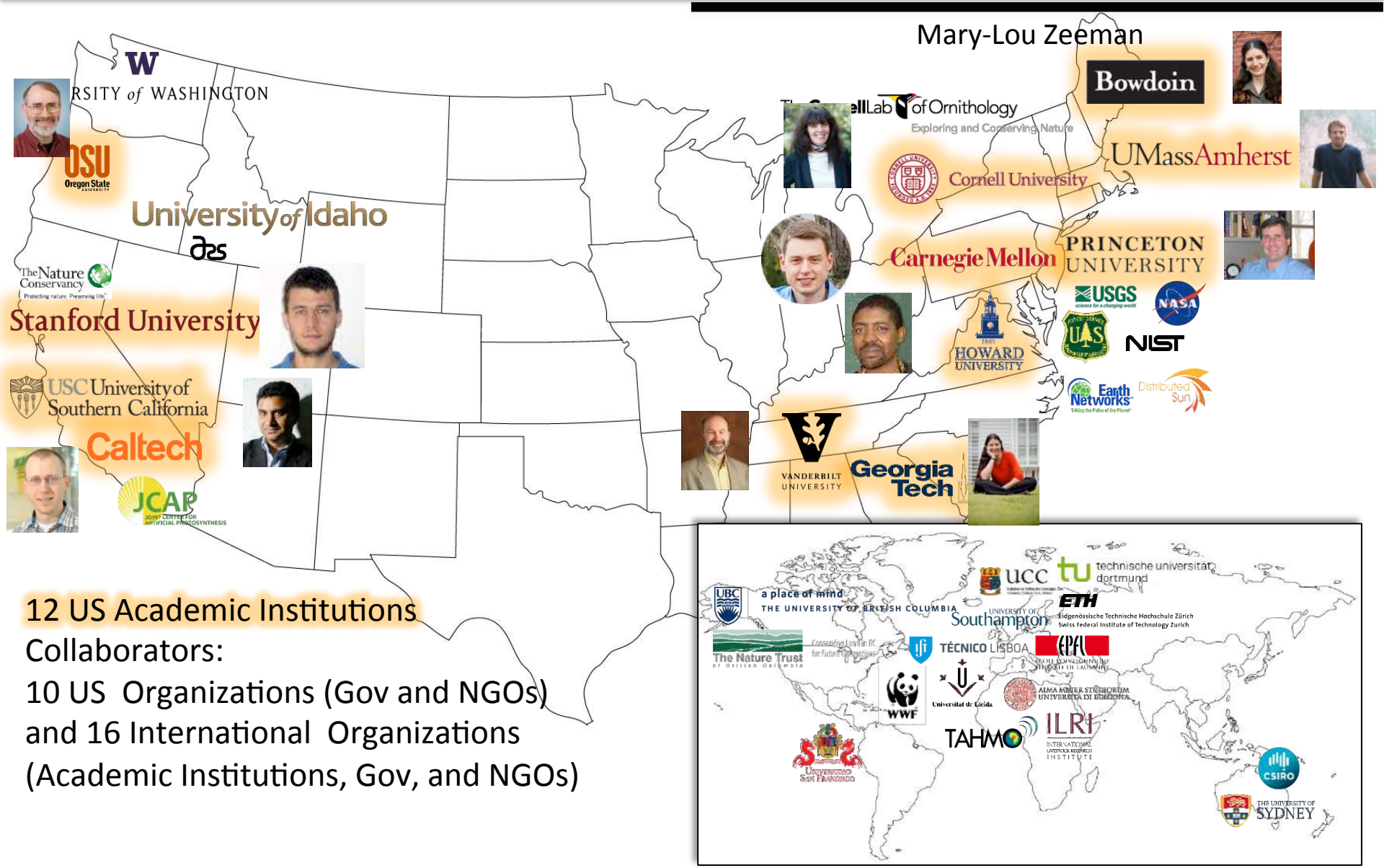
Expeditions in Computing (CISE)

CompSustNet
Large Research Network for
Expanding the Horizons of
Computational Sustainability

CompSustNet:
Large-scale Research Computational Sustainability Network

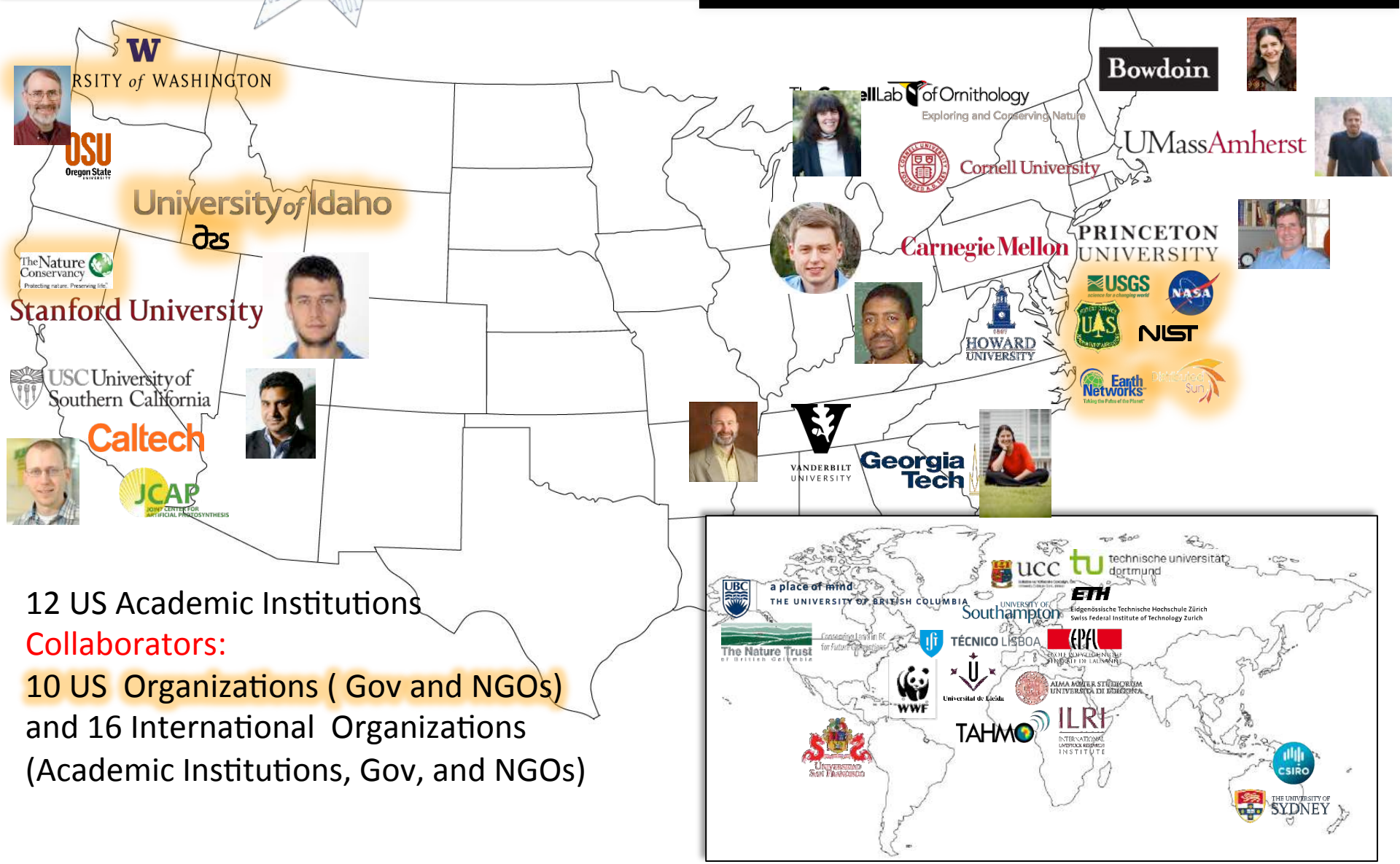
Expeditions in Computing (CISE)

Mary-Lou Zeeman



12 US Academic Institutions
Collaborators:
10 US Organizations (Gov and NGOs)
and 16 International Organizations
(Academic Institutions, Gov, and NGOs)

CompSustNet: Large-scale Research Network



12 US Academic Institutions

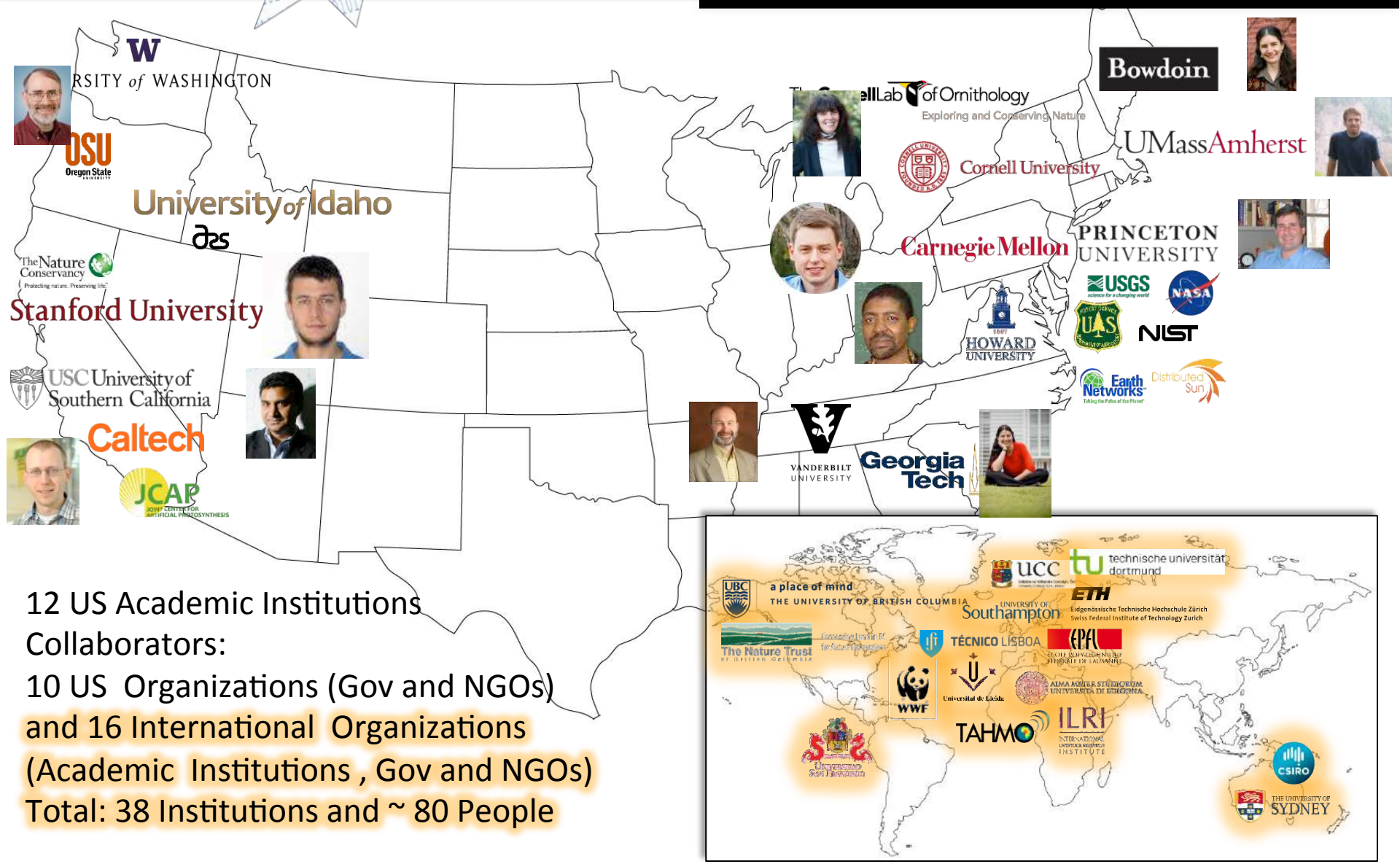
Collaborators:

10 US Organizations (Gov and NGOs)

and 16 International Organizations

(Academic Institutions, Gov, and NGOs)

CompSustNet: Large-scale Research Network



12 US Academic Institutions

Collaborators:

10 US Organizations (Gov and NGOs)

and 16 International Organizations (Academic Institutions , Gov and NGOs)

Total: 38 Institutions and ~ 80 People

Strategically Partnering with Key Centers of Excellence

(providing sustainability knowledge and infrastructure)

The Nature Conservancy

International Livestock Research Institute (Kenya)

Join Center for Artificial Photosynthesis

Secretary of Environment Quito, Ecuador

Trans African Hydro Meteorological Observatory

The National Institute of Standards and Technology Materials Science

NASA Remote Sensing

US Geological Survey

Cornell High Energy Synchrotron Source

Atkinson Center for a Sustainable Future

Laboratory for Energy Systems Analysis

CompSustNet Themes

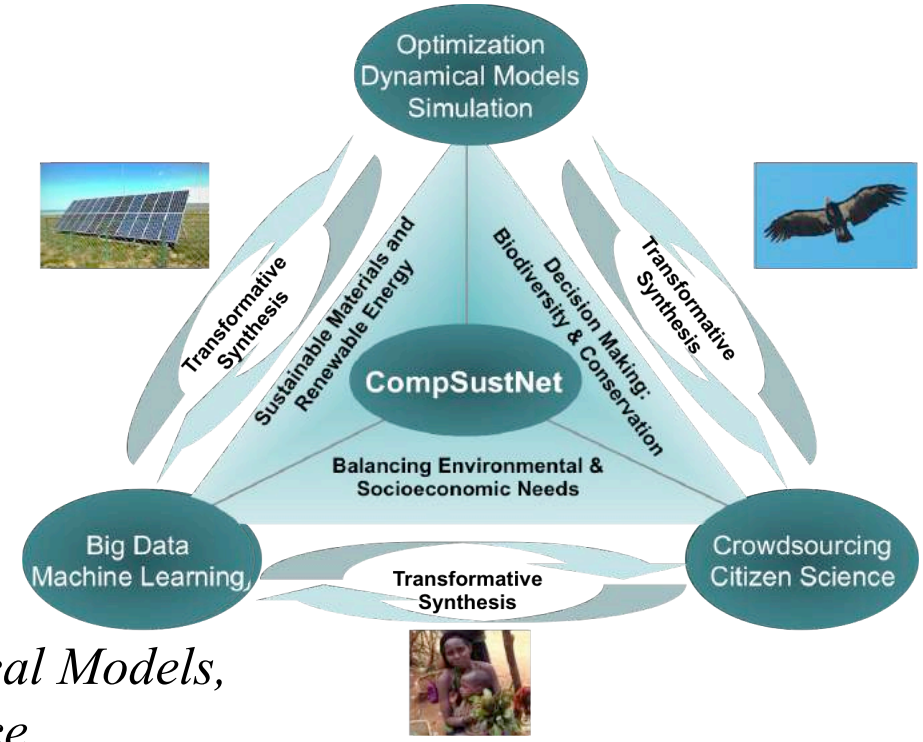
Over 20 Novel Interdisciplinary Research Projects (IRPs)

Core sustainability themes:

(1) Biodiversity and Conservation,

(2) Balancing Environmental and Socio-economic Needs, and

(3) Energy and Renewable Resources.



Main computational thrusts:

(1) Big data and Machine Learning,

(2) Simulation, Optimization, and Dynamical Models,

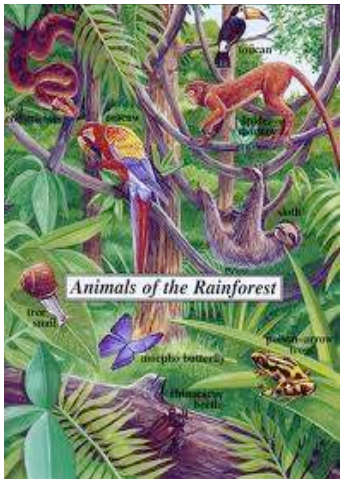
(3) Multi-Agent Systems and Citizen Science.

Wide range of sustainability applications:
To build a ***coherent and cohesive computer science sub-field*** it's paramount to **go deep** in terms of **Cross-Cutting Core Computational Problems** and apply them to different sustainability areas.

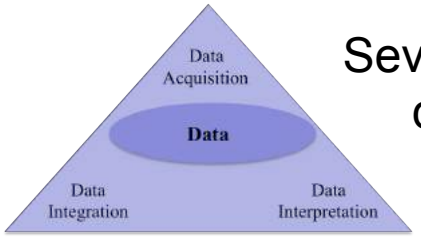
Outline

- Sample of Computational Sustainability Interdisciplinary Research Projects (IRPs)
- Cross-Cutting Computational Themes
- Community Building
- Conclusions

Biodiversity or Biological Diversity



Fundamental question in biodiversity research:
How different species are distributed across landscapes over time.



Several computational data challenges

Sensing biodiversity: A variety of sensors

Global-regional coverage

Satellite

Satellite

Landsat: https://www.youtube.com/watch?v=Ezn1ne2Fj6Y

Regional-local coverage

Aircraft

UAV

Citizen scientist with cell phone

Drone

Wildlife officer with receiving antenna

Transmitting collar

Camera trap

Sound recorder

Collecting envi

Citizen Scientists:
Very sophisticated sensor;
Play a key role in sustainability applications

Sensing Biodiversity, *Science* **346**, 301 (2014); Woody Turner

eBird: Citizen Science at the Cornell Lab. Of Ornithology

- ***Increase scientific knowledge***

  Gather meaningful data to answer large-scale research questions

- ***Increase conservation action***

  Apply results to science-based conservation efforts

- ***Increase scientific literacy***

  Enable participants to experience the process of scientific investigation and develop problem-solving skills

The Citizen Science project at the Lab of Ornithology at Cornell empowers everyone interested in birds to contribute to research by submitting bird observations to the eBird webportal.

Citizen Science for Bird Conservation

300K+ volunteer birders	300M+ bird observations	22M+ **hours of field work** (2500+years)

eBird

CORNELL LAB of ORNITHOLOGY

Bird Distributions

Machine Learning and Citizen Science

eBird

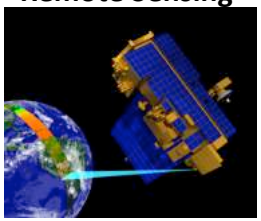
Bird Observations

300K+ volunteer birders	300M+ bird observations	22M+ **hours of field work** (2500+years)

Land Cover

Weather

Remote Sensing



Environmental Data

TeraGrid

80,000+ CPU Hours (~ 10 Years!!!)

Adaptive Spatio-Temporal Machine Learning Models and Algorithms

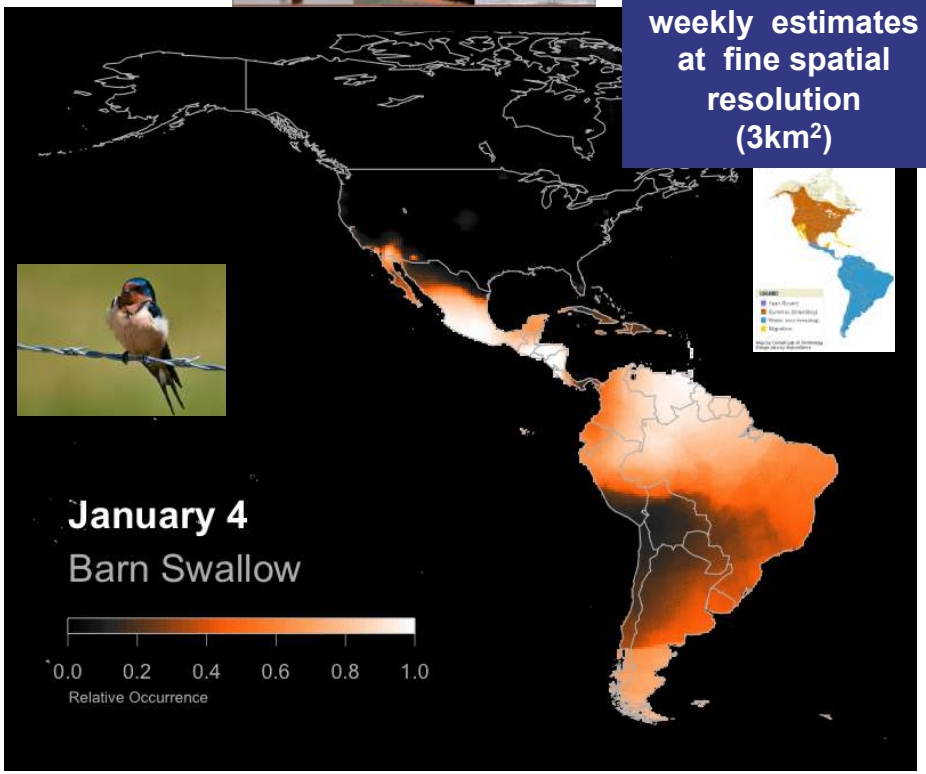
Relate environmental predictors to observed patterns of occurrences and absences

State of the Birds Report

(officially released by Secretary of Interior)

Novel Approaches To Conservation Based on eBird Models

Distribution Models for 400+ species with weekly estimates at fine spatial resolution (3km^2)



Patterns of occurrence of the Barn Swallow for different months of the year Source: Daniel Fink

1st Time

Hemisphere Scale Bird Distribution Models, Revealing, at a fine resolution, Species' Habitat Preferences

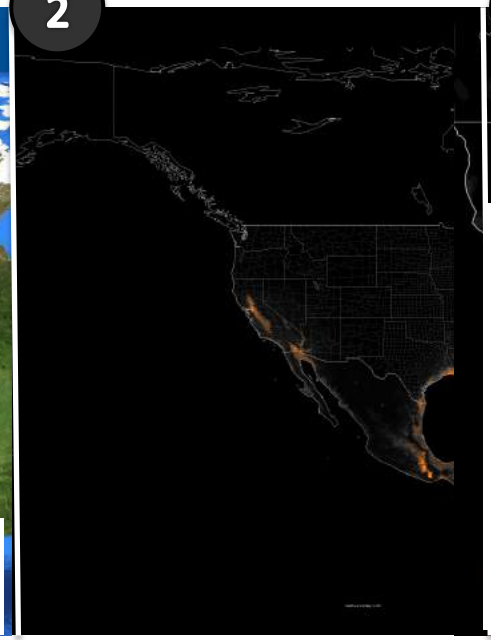
Dynamic High-Precision Conservation
Bird Returns

Protecting Migratory WaterBirds in California Against Drought by paying farmers to keep water in rice fields during the bird migration

1

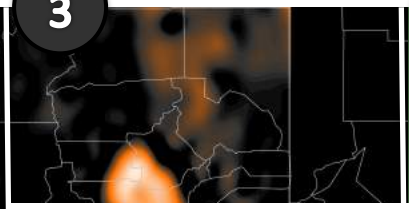
Pacific Migration Flyway

2



eBird Models

3



Target Estimates

4

Farmers Submit Bids

Reverse Auction Bid Selection

Combinatorial Reverse Auctions

Using big data, machine learning techniques, and a market-based approach
The Nature Conservancy has generated
over 20,000 acres of additional habitat for waterbirds in California

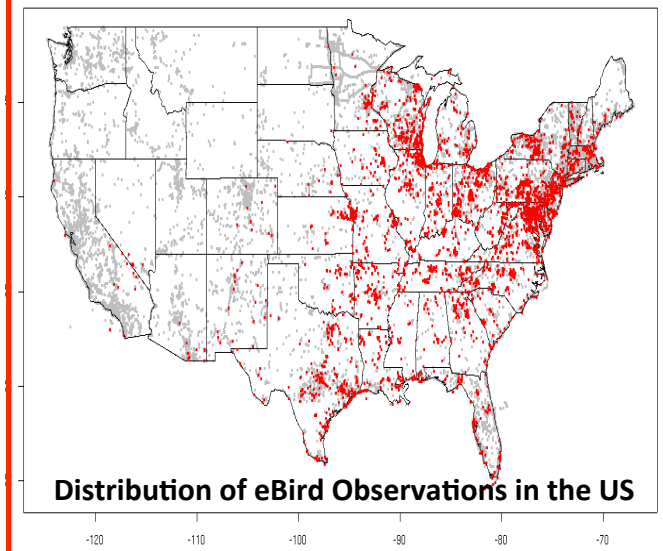
Sacramento Valley, CA

Farmers submit bids to keep the rice fields flooded during short periods of bird migration in California.

Radically novel, dynamic, way of doing bird conservation.
Only possible because of advanced computational methods.

Avicaching:
A Two Stage Game for Bias Reduction in Citizen Science

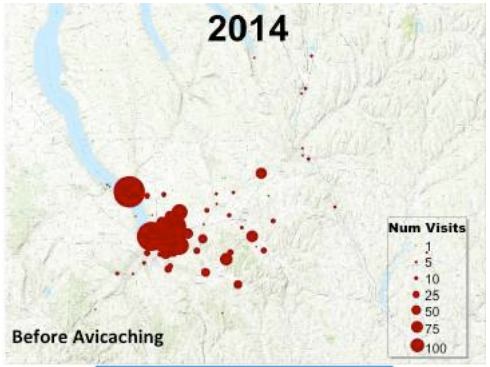
Data Bias Problem

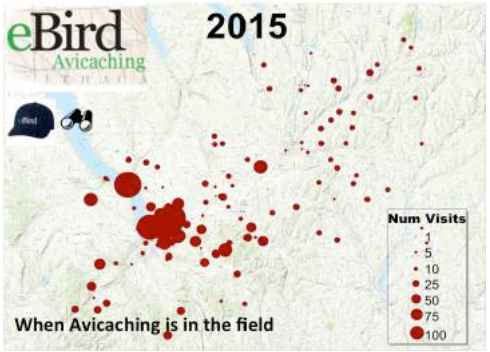


Prevalent problem in citizen science
Collected data are often aligned with the participants' preferences rather than scientific objectives.

How to incentivize Citizens to visit under-sampled areas?

Principal-Agent Framework





Very Successful in Two US Counties (19% shift to undersampled areas)

Field: Pilot Program

eBird Avicaching

Incentivize eBirders to visit undersampled locations.

Incentives:

- Avicaching points,
- leaderboards
- Lotteries (e.g. binoculars.)

Socio-ecological corridor for Andean Bear in Ecuador

REU

Planning Conservation Decisions in Ecuador
Integrating Socio-Economic Factors

Secretaría de Ambiente

Ecuador: Top Biodiversity Hotspots in the World

Photo: Santiago Molina

Andean Bear:
Umbrella species

Habitat needs represent minimal requirements for numerous co-occurring species

Challenge: SocioBioLinks Corridor Design

→Protecting biodiversity
Population of Andean bears and other species
→ Improving the lives of local communities
(level of welfare/economic stability of local residents
→ecotourism, cooperatives, other activities)

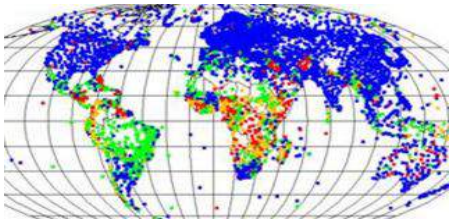
Examples in of
Other Interdisciplinary Projects

Poverty Mitigation in Africa:
Big Data for Africa

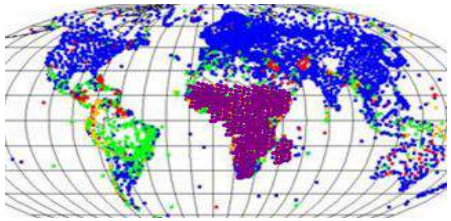
Improving Weather, Rangeland, and Forage Maps in Africa
to protect farmers and herders

Deployment of 20,000 Low-cost Weather-Stations in African Schools

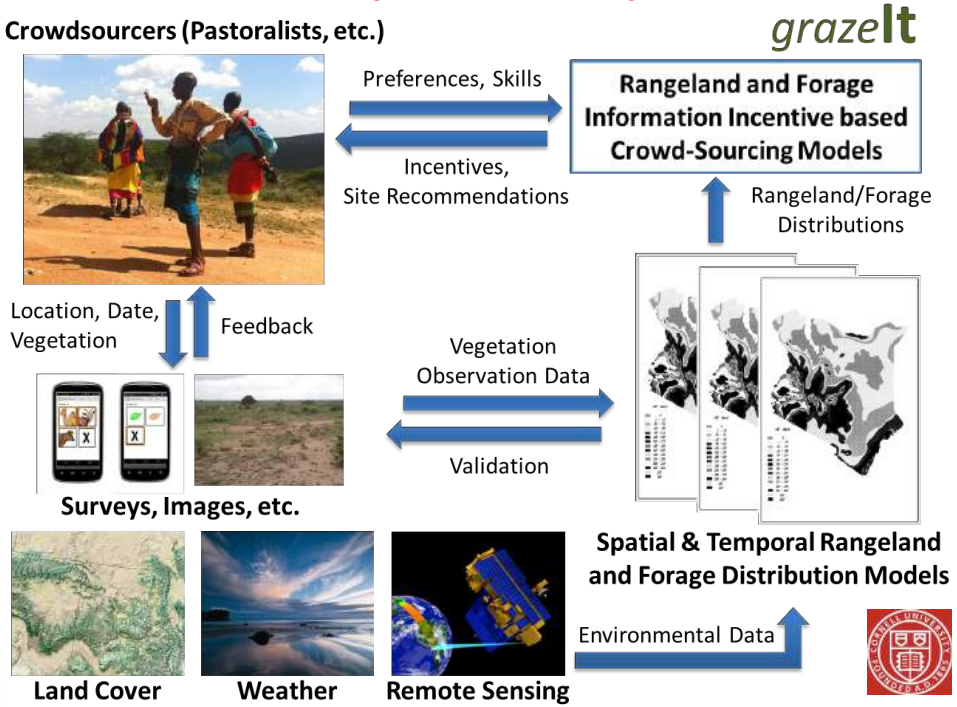
Weather stations in Africa (blue points in map)



Projected Weather stations in Africa (purple points in map)



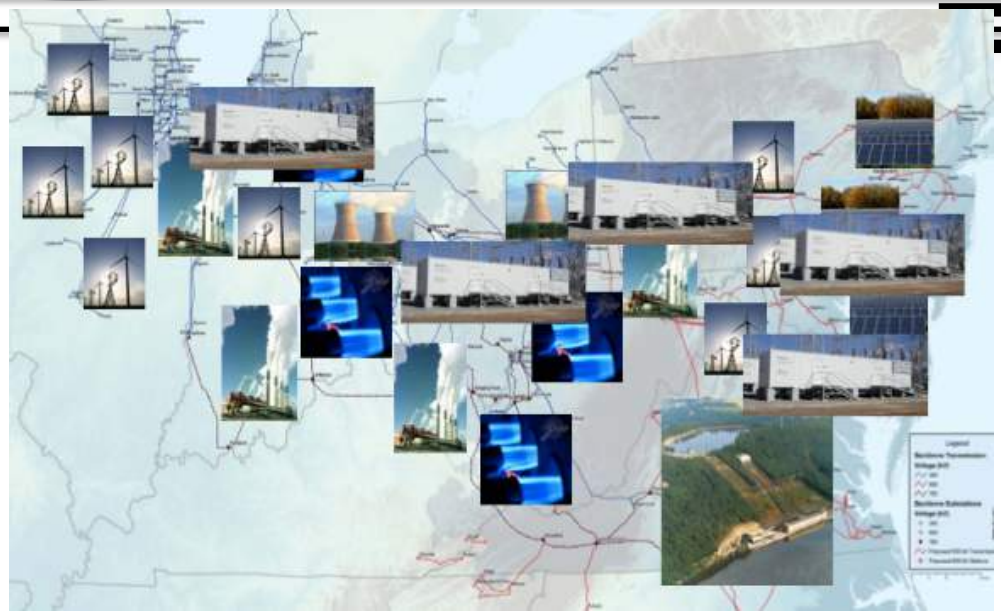
Herders Submit Vegetation Images and Surveys with Smartphones



3 month Pilot project: → 200,000+ surveys

Africa is very poorly sensed
(limited environmental data, vegetation maps, only a few reliable weather stations)

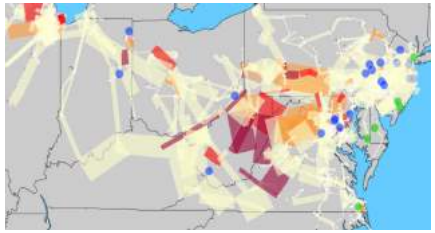
Dynamic Energy Storage and Economic Dispatch in the Presence of Renewables for the SMART Grid



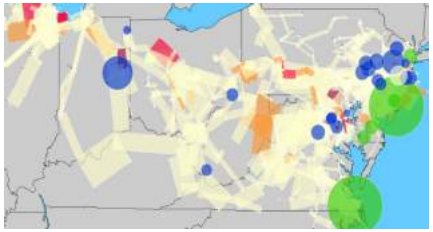
Smart Grid with Different Types of Generation
(Wind; Solar; Hydroelectric; Nuclear, Coal; Gas Turbines) +Storage

Renewables are non-dispatchable

→ Modeling uncertainty
→ Robust plans combining renewables and traditional power resources



No Storage: Congestion (red - power flowing at high congestion levels)

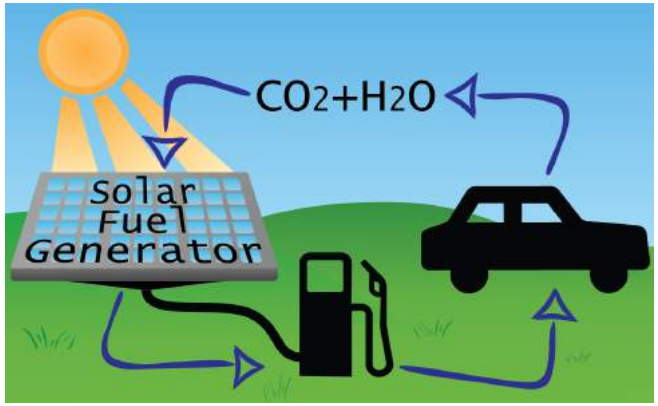


Optimizing storage: No congestion (green circles -storage near wind farms; and blue circles (circles near load points)

Multiple times scales

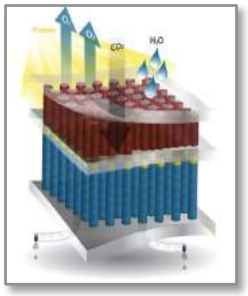
Due to, e.g.
→ Different types of generation
→Fine grained frequency regulations (every few seconds),
→Energy shifting (spanning hours) →Peak capacity charges (spanning year).

Accelerating Materials Discovery for Sustainable Energy



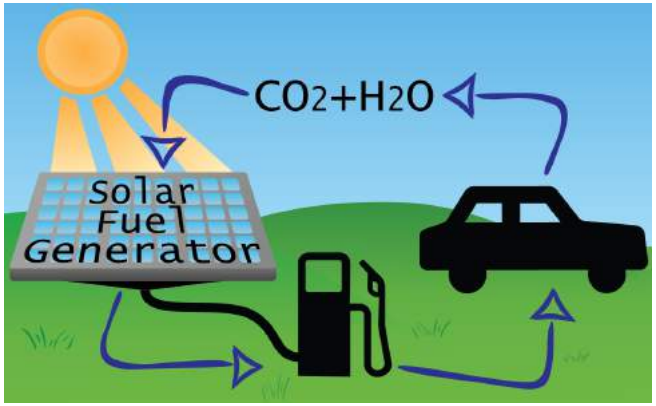
Accelerating the Development of Solar Fuel Generators

"… existing energy approaches […] *will not be enough to secure our energy future. Meeting the challenge will require […] scientific breakthroughs in new materials….*"
Report of Dept. of Energy [2010]



Solar fuel Cell

Solar fuels can be substantially produced and stored



Picture by John Gregoire

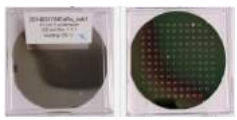
Renewable transportation fuels can be synthesized using solar energy, water and carbon dioxide

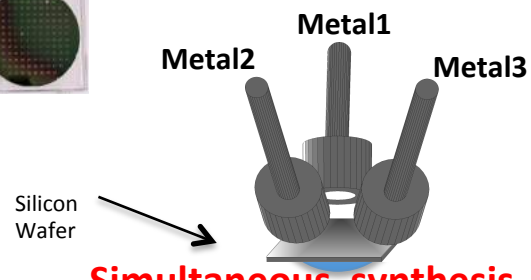
JCAP Caltech
JOINT CENTER FOR ARTIFICIAL PHOTOSYNTHESIS

CHESS
Cornell High-Energy Synchrotron

Discovering better catalysts for solar fuel generators

Co-sputtering
(similar to atomic spray painting)

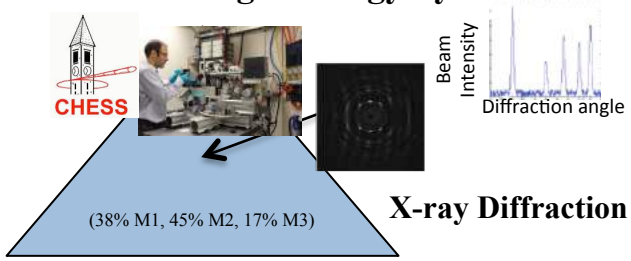




Simultaneous synthesis
of thousands of materials

High-Throughput
Materials Discovery

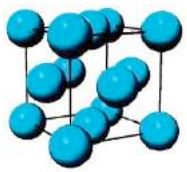
Cornell High-Energy Synchrotron



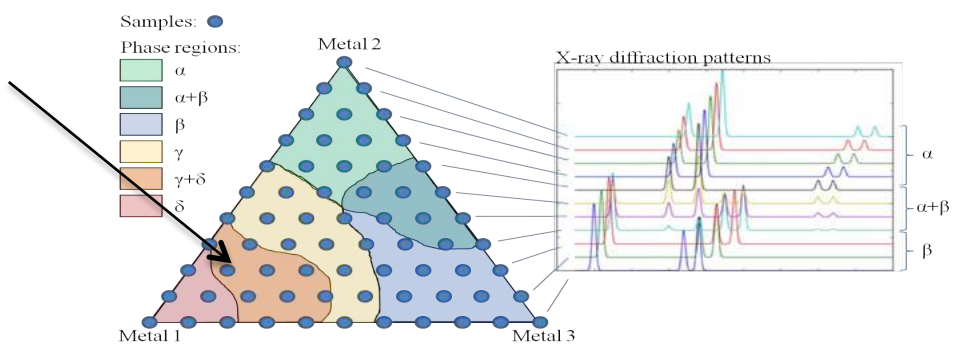
X-ray Diffraction

Rapid characterization
of thousands of materials

How to determine the crystal structure
of the materials, based on the X-ray diffraction patterns?



FCC Crystal Structure



Goal:
Achieve High-Throughput Crystal Structure Identification

Crystal Structure Map
Problem:
Infer the crystal structures of the materials from the X-ray diffraction patterns

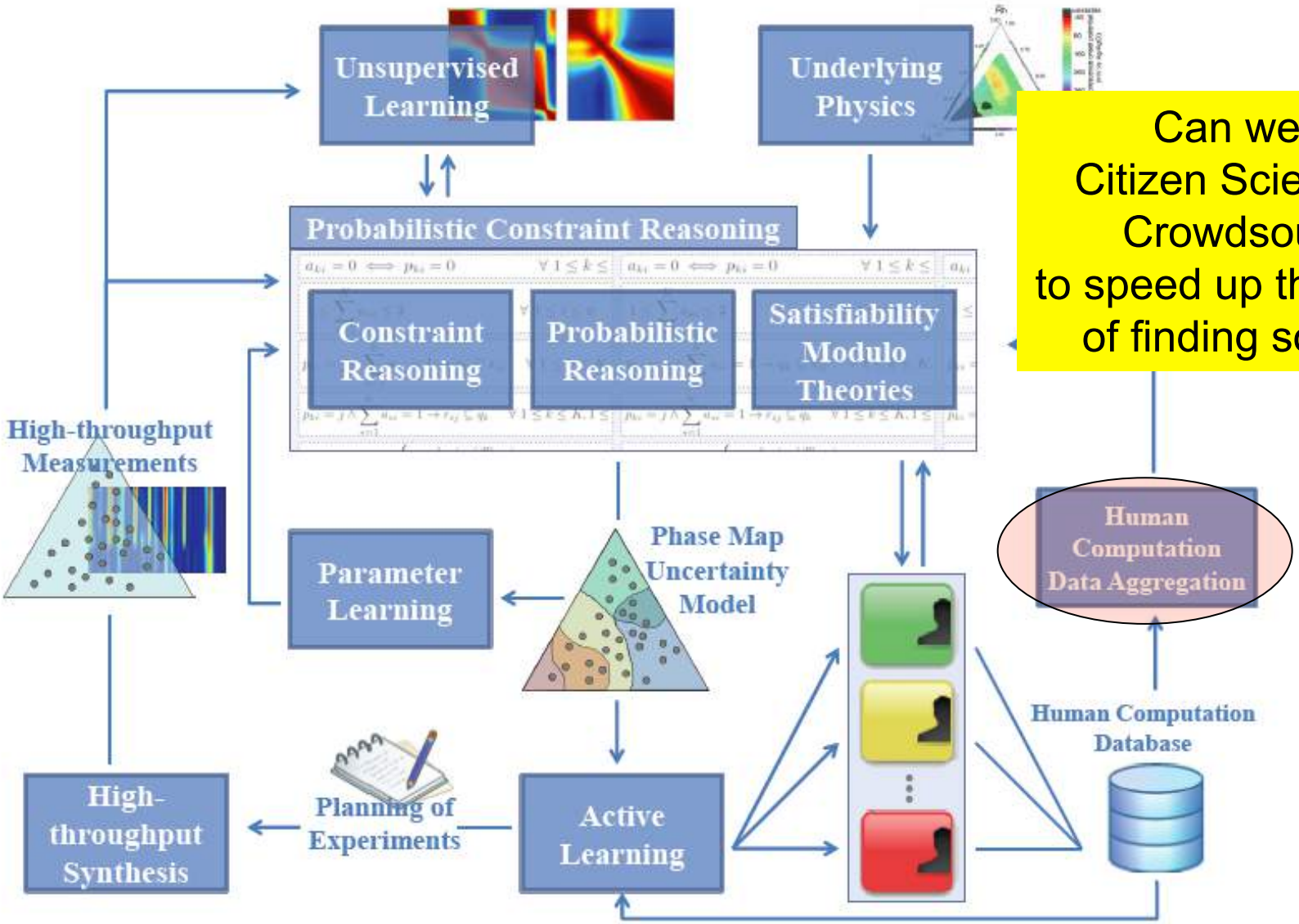
(currently a largely manual process, requiring expert knowledge)

Difficulty: Often X-ray diffraction patterns correspond to a mixture of crystal structures

Challenging to un-mix the X-ray diffraction patterns

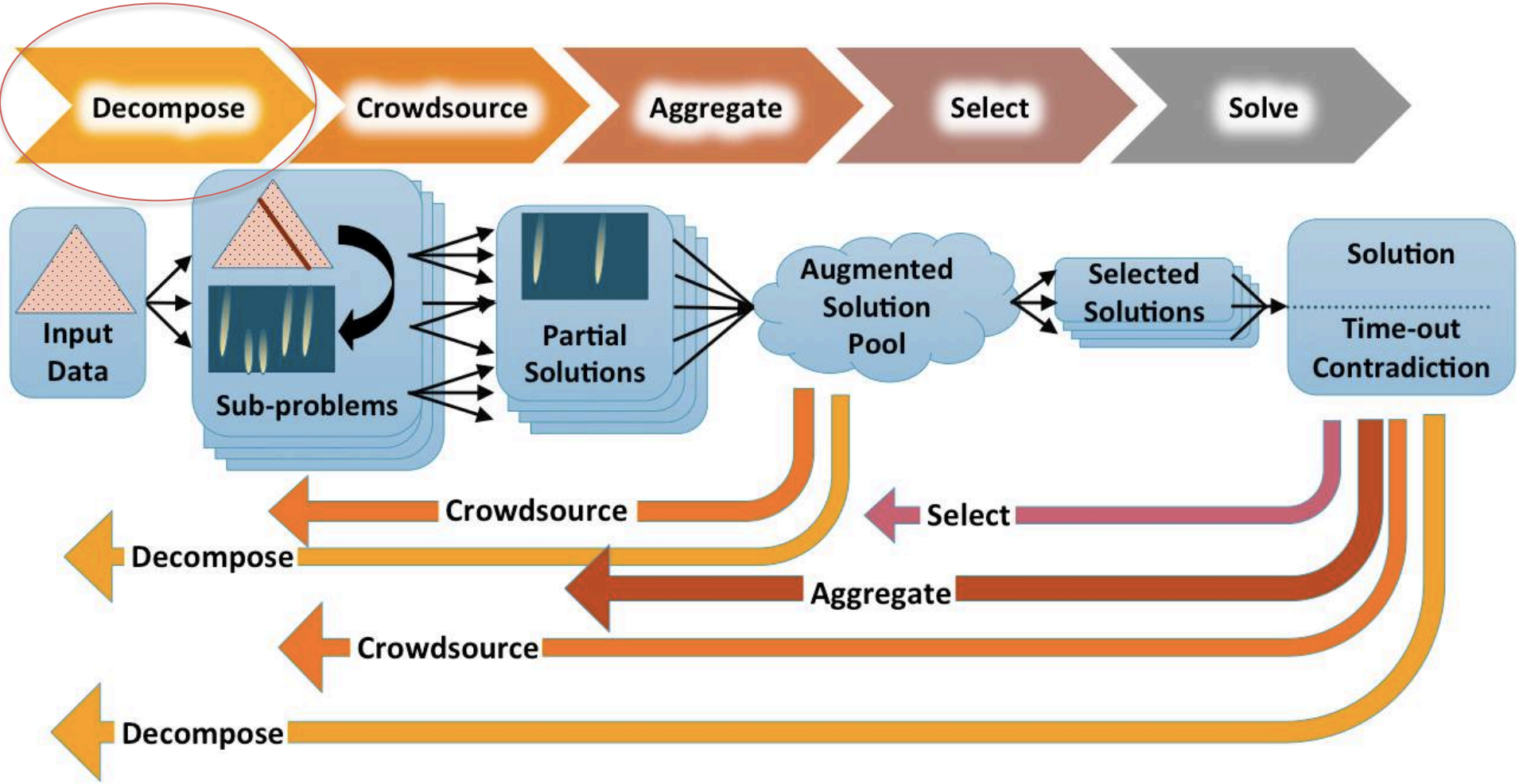
JCAP JOINT CENTER FOR ARTIFICIAL PHOTOSYNTHESIS
Caltech

Combinatorial Reasoning, Learning, and Human Computation



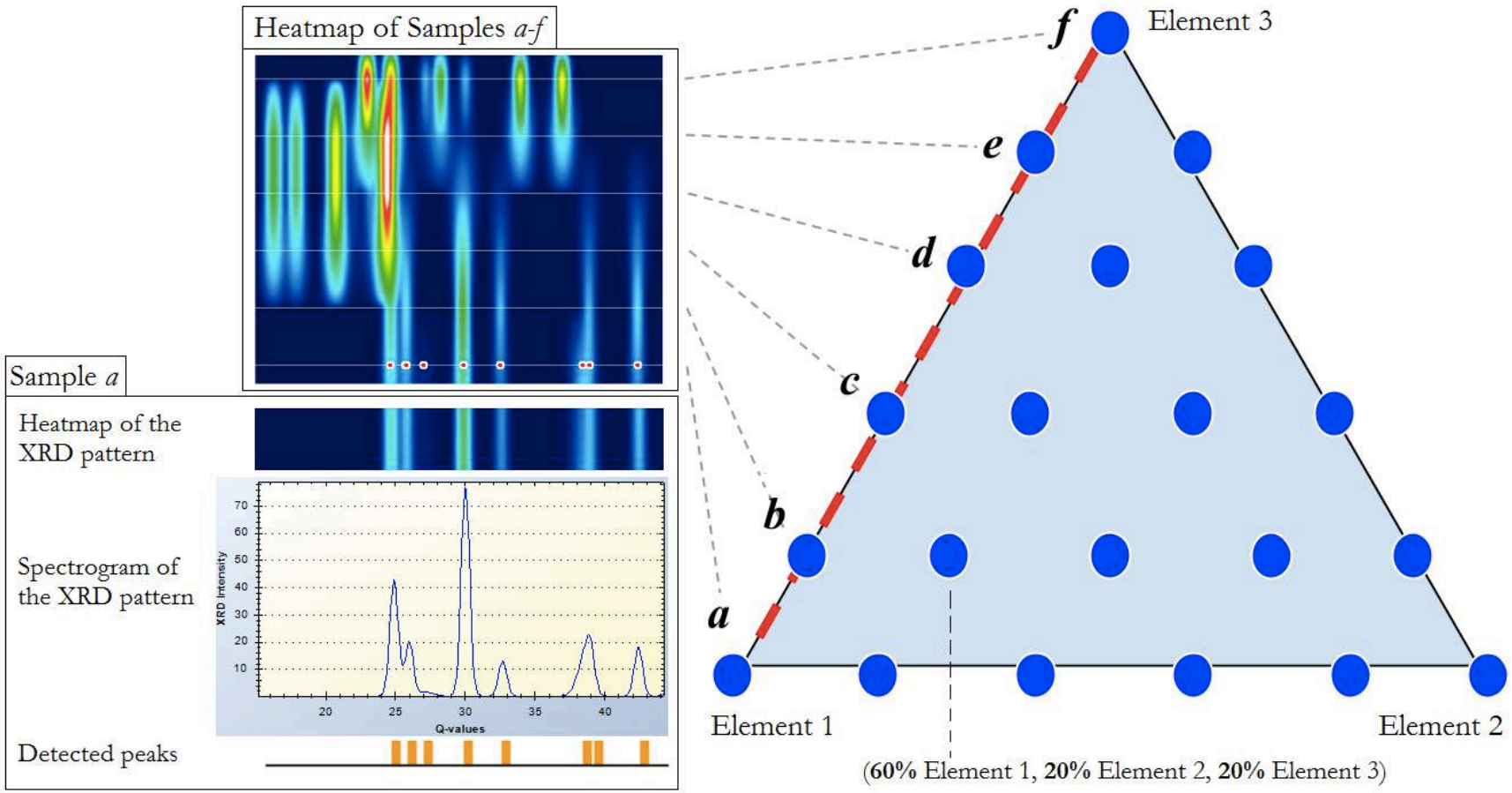
Can we use
Citizen Science and
Crowdsourcing
to speed up the process
of finding solutions?

Human Computation Framework for Boosting Computational Solvers

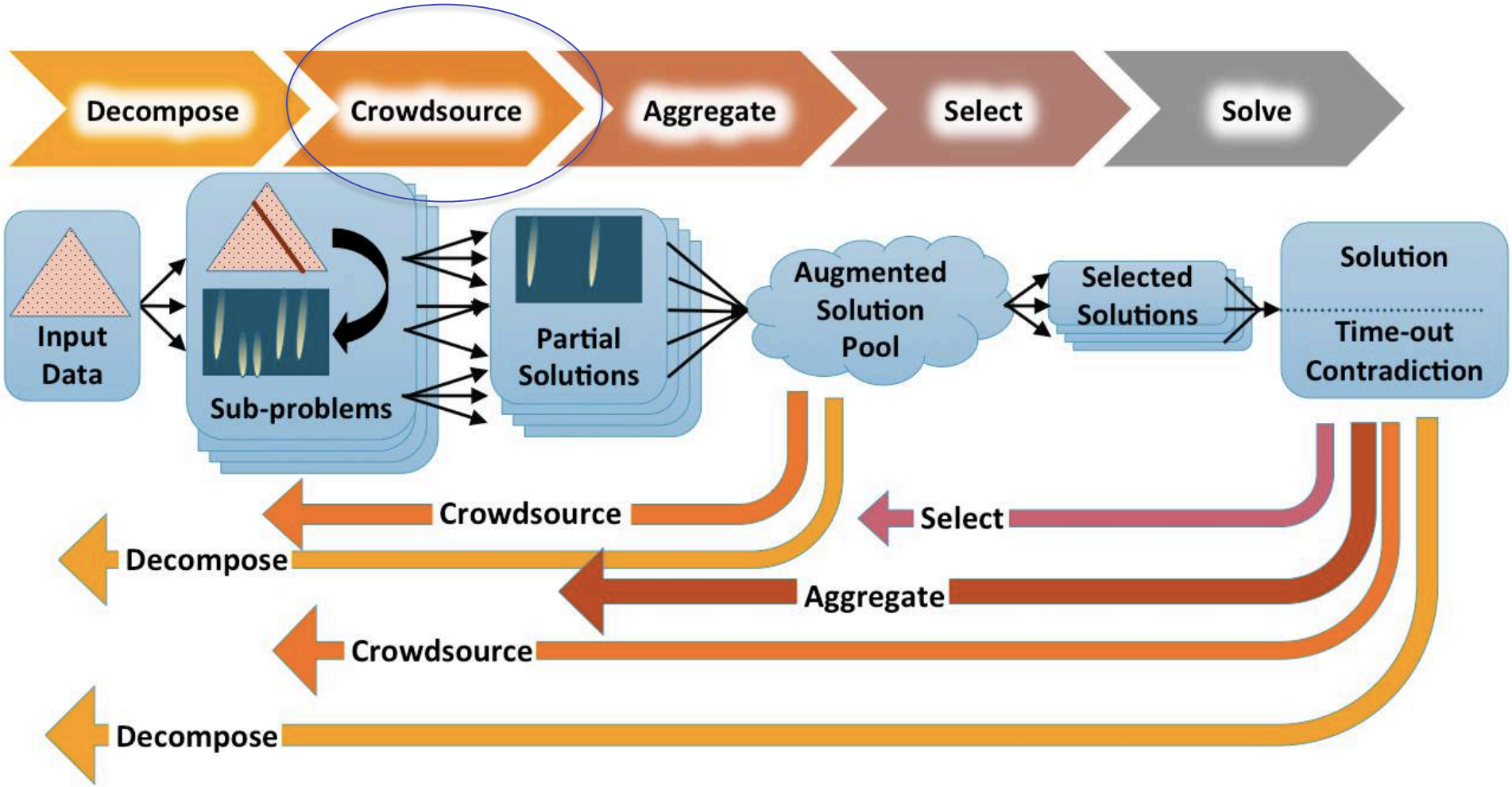


Ronan Le Bras, Yexiang Xue, Richard Bernstein, Carla P. Gomes, Bart Selman, HComp14

Decomposition:
Slices of Sample Points
Different Visualizations of X-Ray Diffraction Patterns

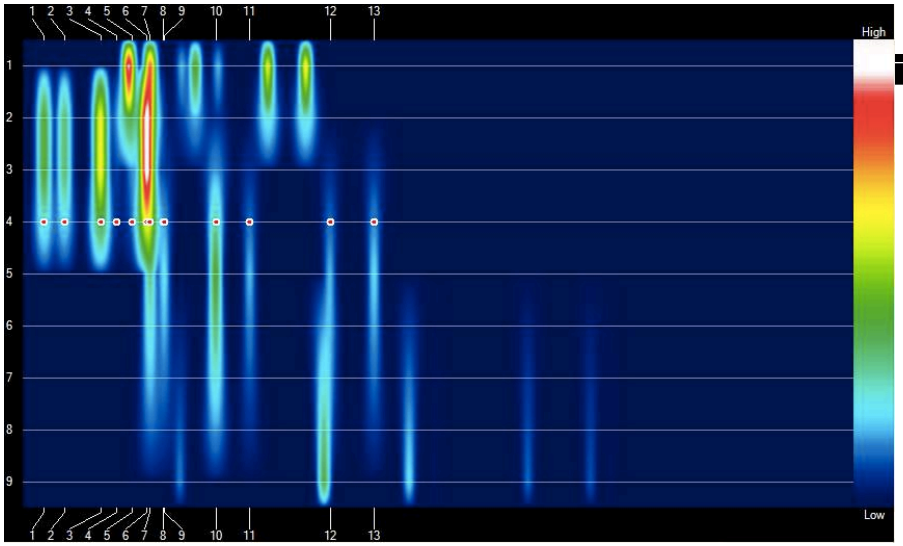


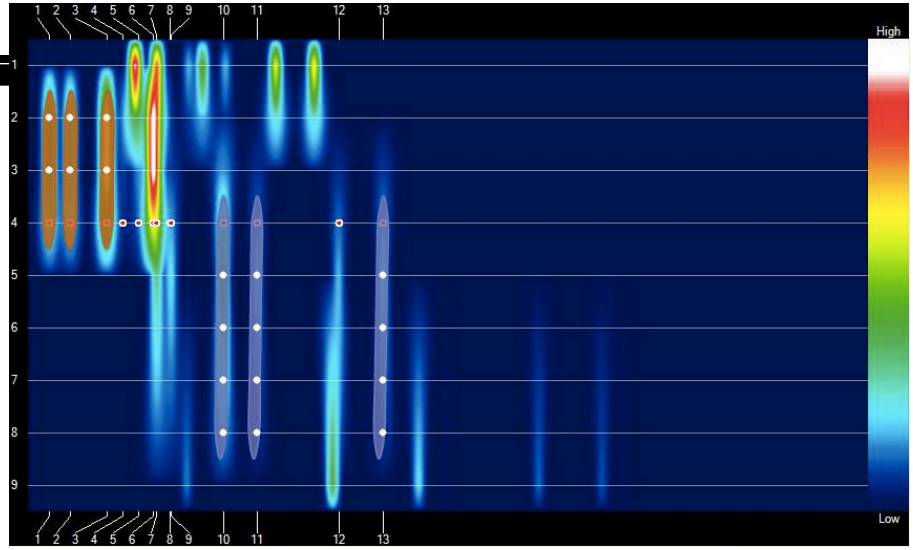
Human Computation Framework for Boosting Combinatorial Solvers



Ronan Le Bras, Yexiang Xue, Richard Bernstein, Carla P. Gomes, Bart Selman, HComp14

U-DiscoverIt:
Citizen-Science and Crowdsourcing: General Public





Human Intelligent Task (HIT)

A worker is asked to identify patterns of similar vertical lines that intersect with sample 4 (whose detected intensity peaks are marked with red dots).

Completed Task

Orange pattern -3 vertical lines.
Purple Pattern - 3 vertical lines.

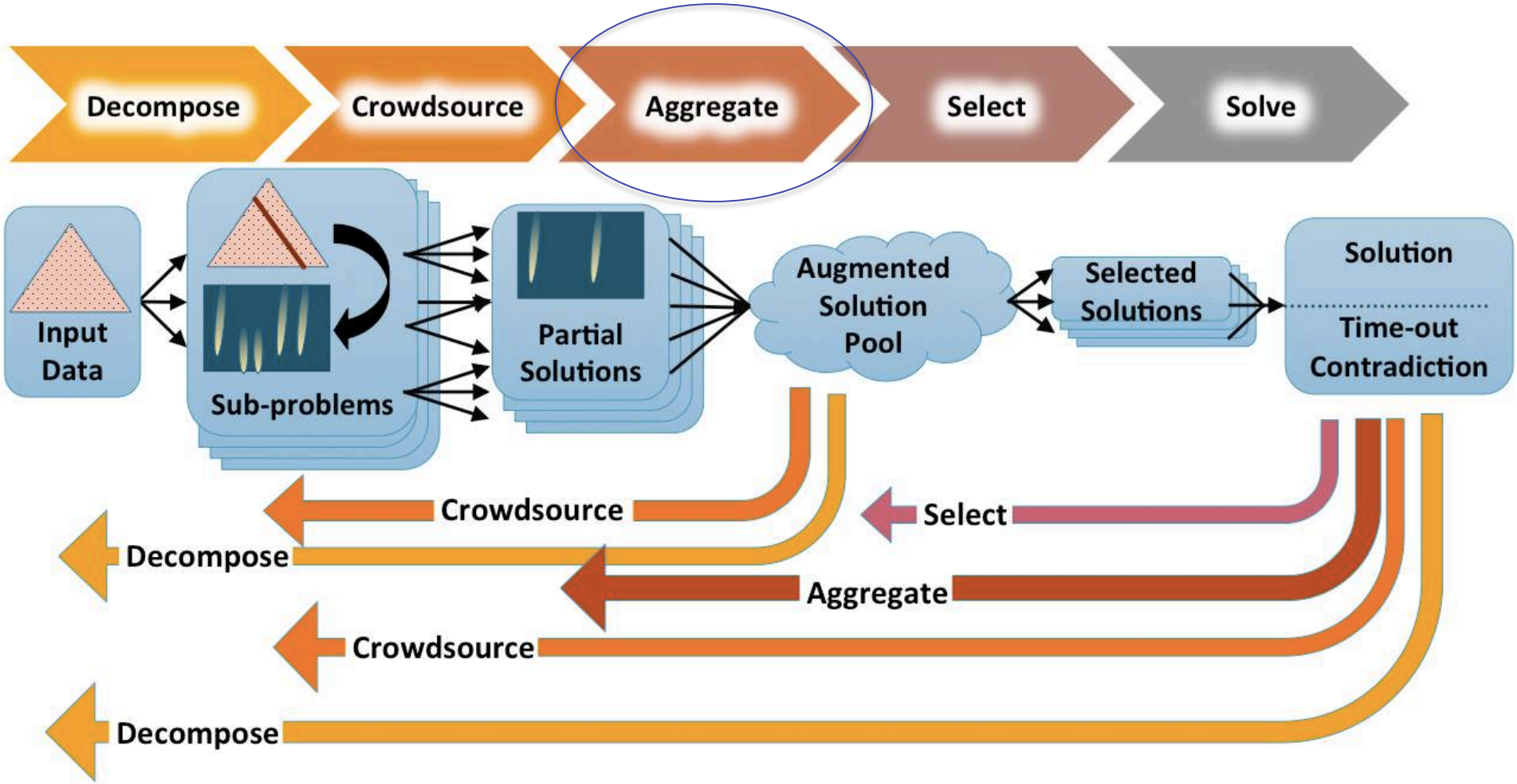
Unmarked - all the others are less clear and because of their ambiguity have been left unmarked, which is the correct decision, since workers are told to be conservative

Cornell Summer College
EXPLORATIONS IN ENGINEERING
ENGRG 1060
~ 90 High-School students

http://www.udiscover.it

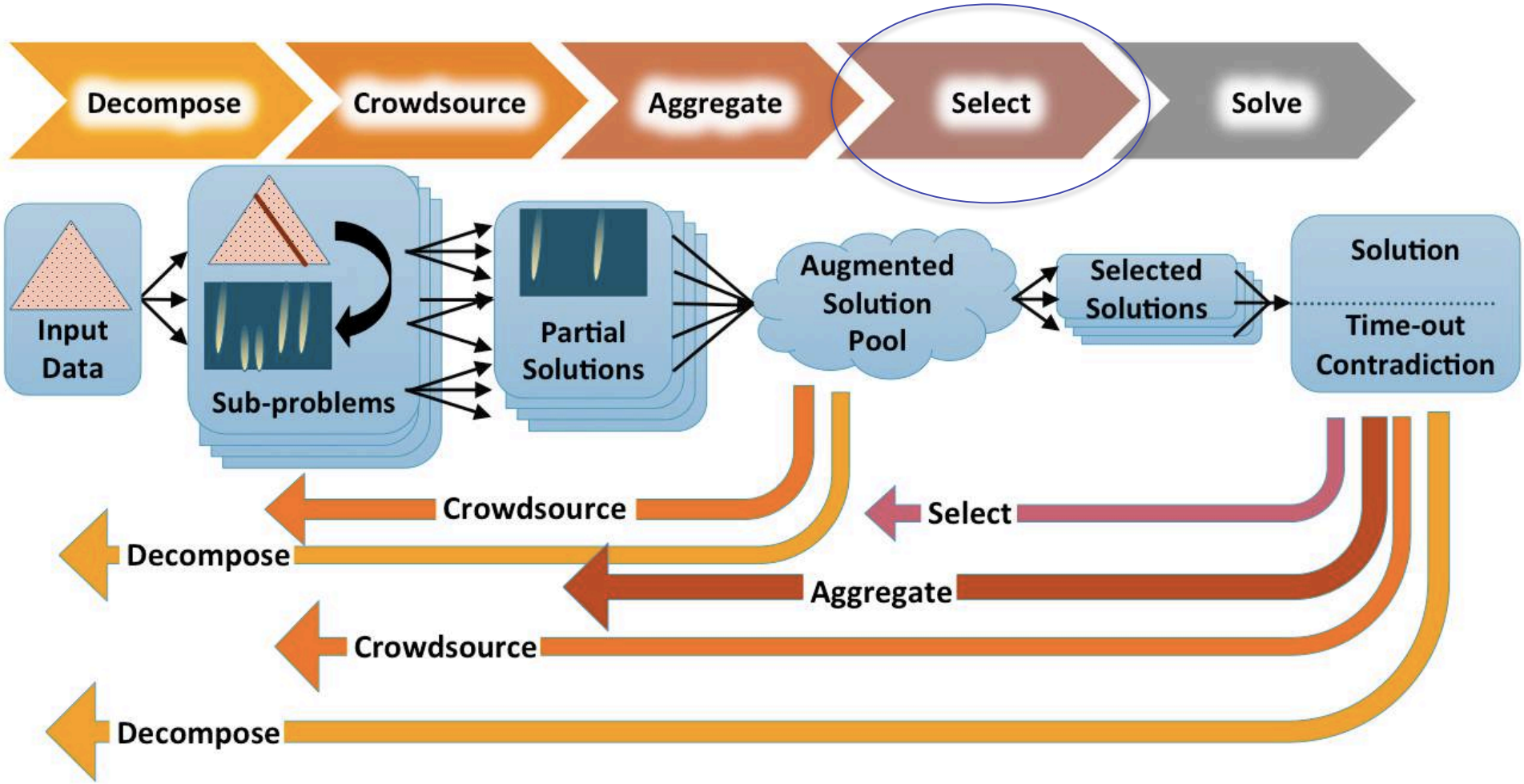
Institute for Computational Sustainability ICS

Human Computation Framework for Boosting Combinatorial Solvers



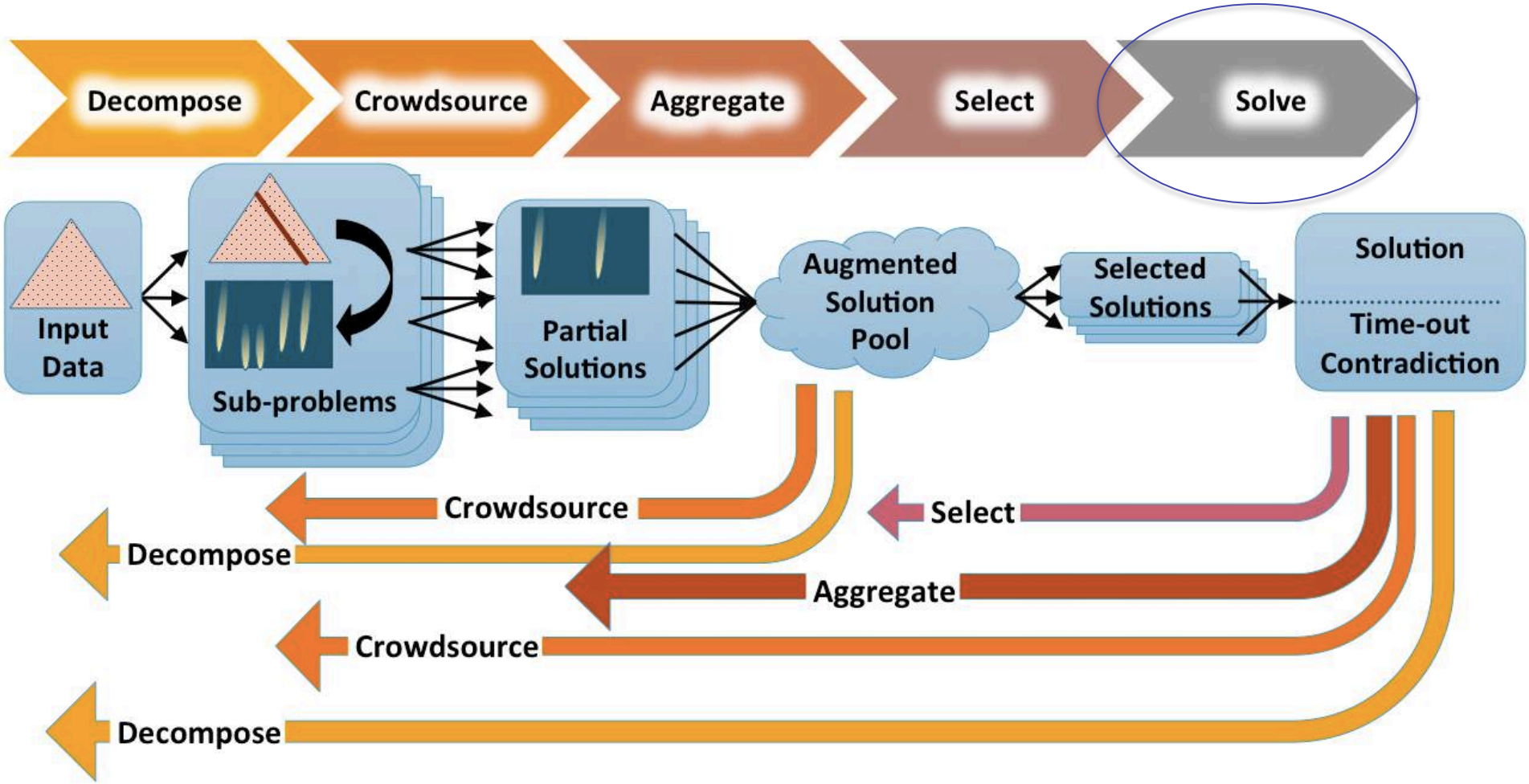
Ronan Le Bras, Yexiang Xue, Richard Bernstein, Carla P. Gomes, Bart Selman, HComp14

Human Computation Framework for Boosting Combinatorial Solvers



Ronan Le Bras, Yexiang Xue, Richard Bernstein, Carla P. Gomes, Bart Selman, HComp14

Human Computation Framework for Boosting Combinatorial Solvers



Ronan Le Bras, Yexiang Xue, Richard Bernstein, Carla P. Gomes, Bart Selman, HComp14

U-DiscoverIt: CrowdSource Mechanical Turk

System	Dataset P	Dataset L^*	Dataset K	#var	#cst	Time w/o user input (s)	Time w/ user input (s)	# assigned var. by user
A/B/C	36	8	4	408	2095	3502	859	19 (4.6%)
A/B/C	60	8	4	624	3369	17345	4377	18 (2.9%)
Al/Li/Fe	15	6	6	267	1009	79	4	6 (2.2%)
Al/Li/Fe	28	6	6	436	1864	346	62	12 (2.7%)
Al/Li/Fe	28	8	6	490	2131	100[illegible]	[illegible]	[illegible]
Al/Li/Fe	28	10	6	526	2309	281[illegible]	[illegible]	[illegible]
Al/Li/Fe	45	7	6	693	3281	188[illegible]	[illegible]	[illegible]
Al/Li/Fe	45	8	6	711	3410	46816	1003	30 (4.2%)

Human Computation:
From ~13 hours to 1000 seconds!

Table 1: Runtime of SMT solver with and without user input. P is the number of sample points, L^* the average number of peaks per phase, K the number of basis patterns, #var the number of variables and #cst the number of constraints.

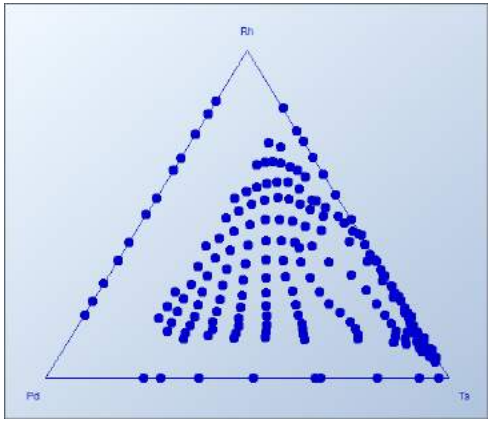
Human input can dramatically speed up the performance of combinatorial optimization methods.

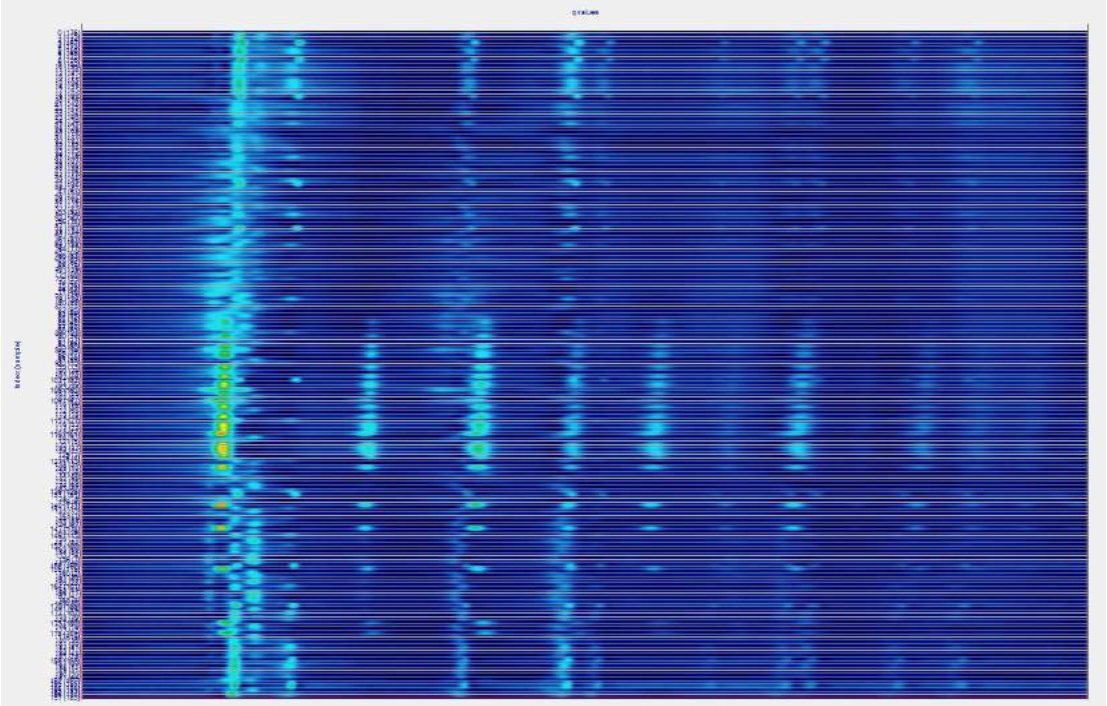
Approaches leverages the complementary strength of human input, providing global insights into problem structure, and the power of combinatorial solvers to exploit complex constraints.

Le Bras, Bernstein, Gomes, Selman and van Dover 2013-2014

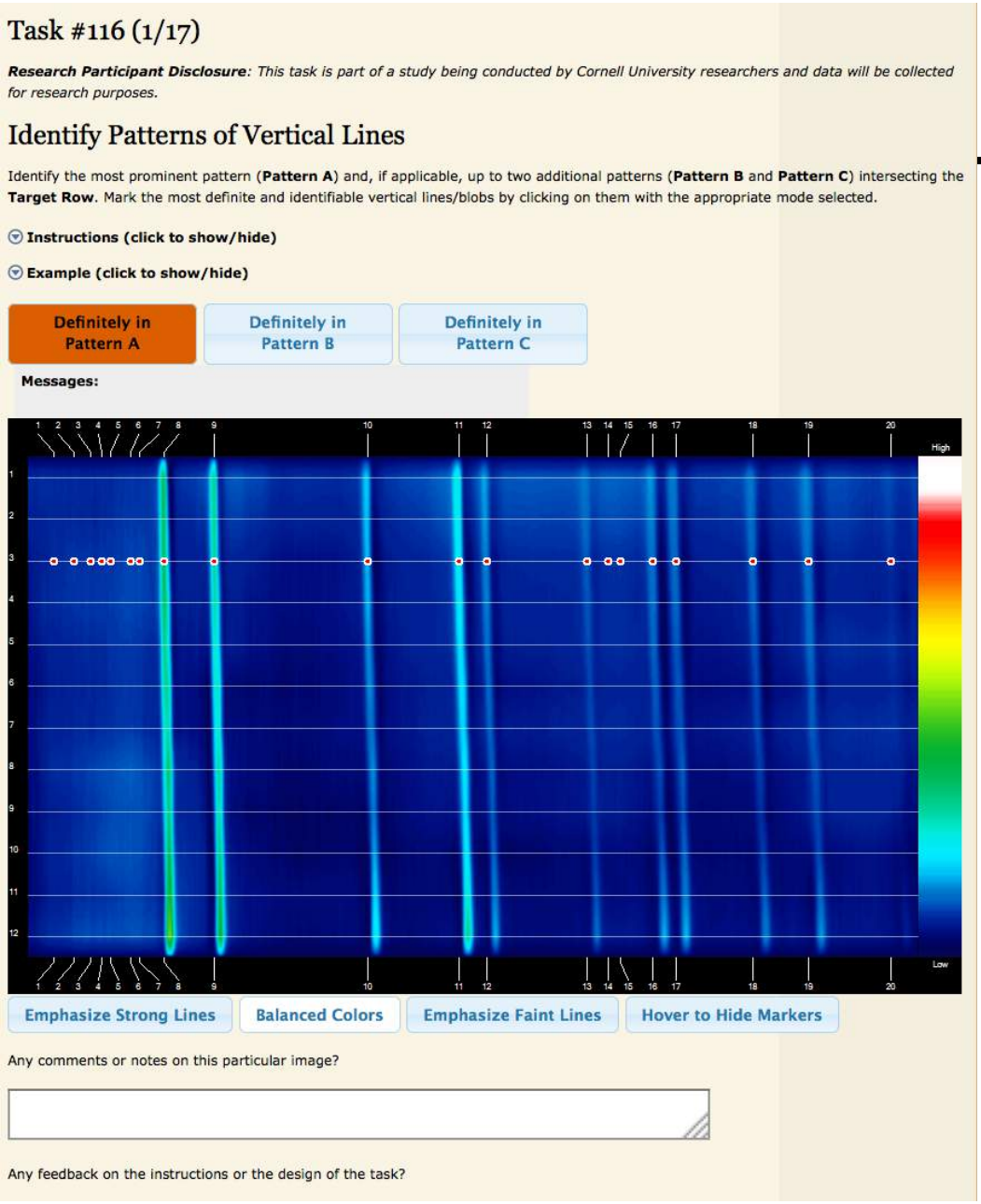
Real System: Pd-Ta-Rh system

198 synchrotron
x-ray diffraction patterns





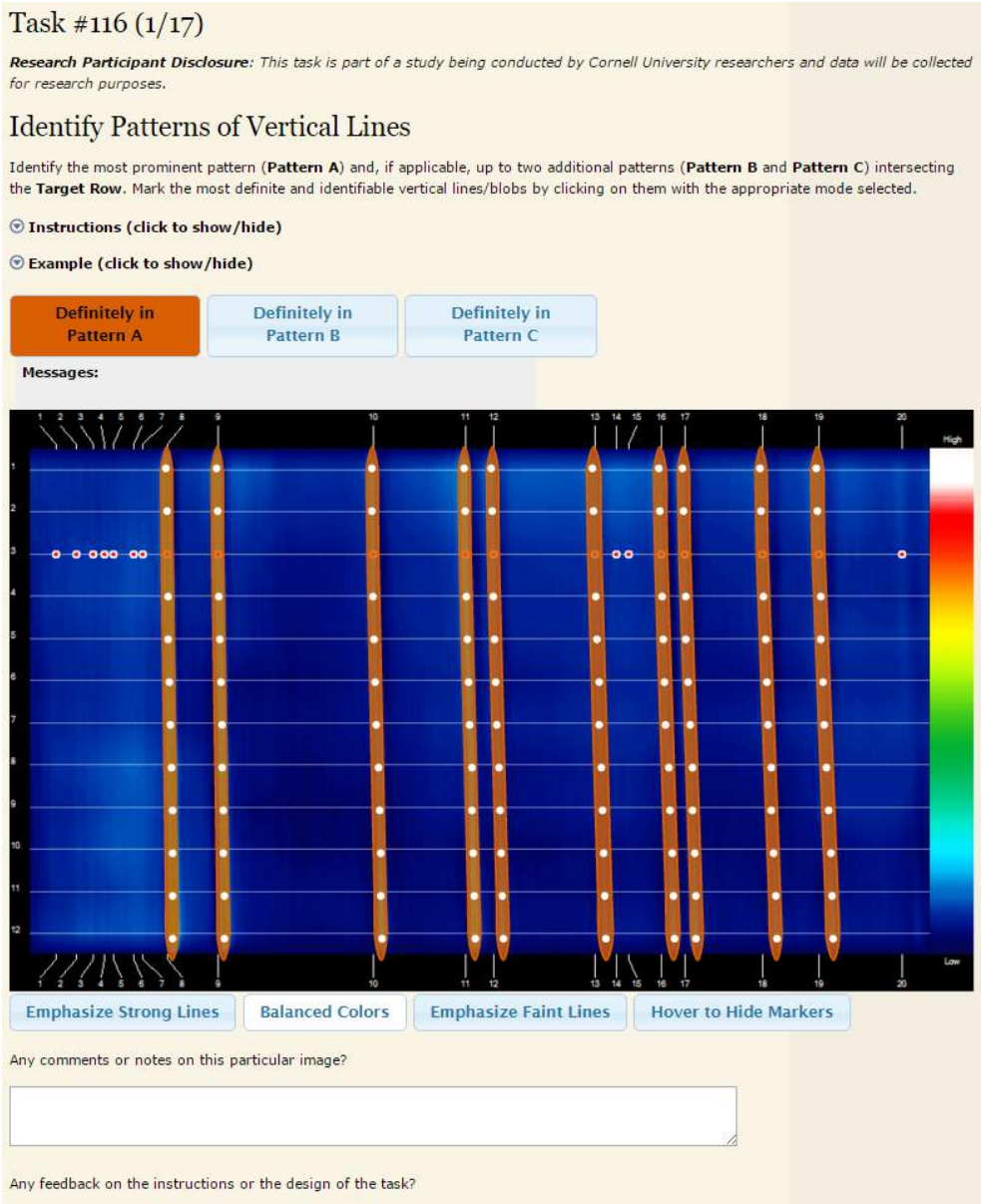
UDiscoverIt: Task #116



Cornell Summer College
EXPLORATIONS IN ENGINEERING
ENGRG 1060
~ 90 students
(Bruce van Dover)

Amazon Mechanical Turk

UDiscoverIt

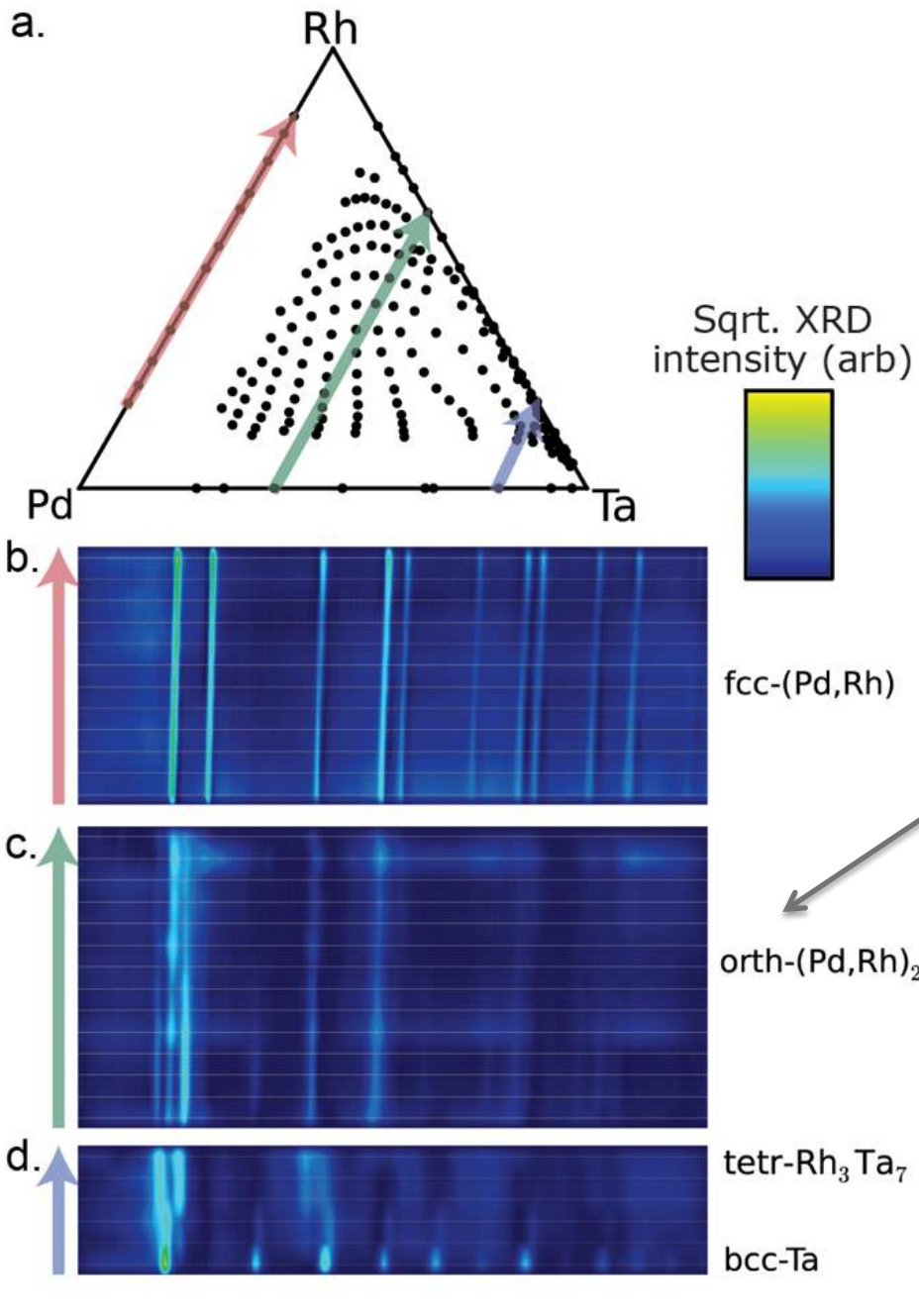


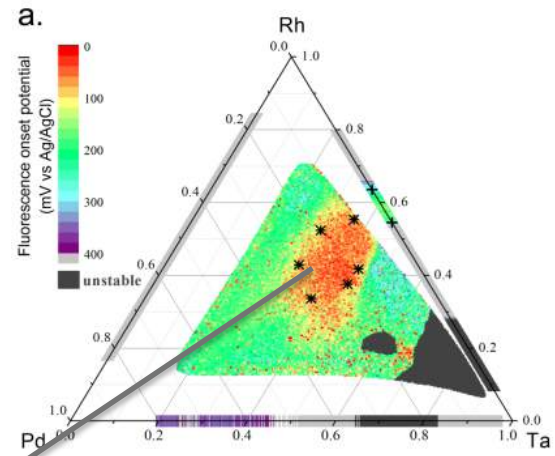
Participants ignored noisy peaks;
Large number aggregated the peaks into single pattern
(8 or more peaks)

Cornell Summer College
EXPLORATIONS IN ENGINEERING
ENGRG 1060
~ 90 students

Amazon Mechanical Turk

Pd-Rh-Ta System: Best solution: Hybrid Human-Machine System





orth-(Pd,Rh)$_2$Ta

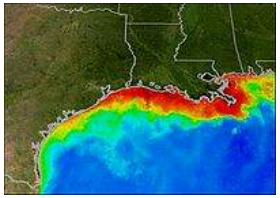
$Pd_{0.14}Rh_{0.40}Ta_{0.46}$ →best known Pt-free catalyst for methanol oxidation

Agriculture

Nitrogen Based Fertilizers

Nitrogen Cycle and Fertilizers

Nitrogen Based Fertilizers



Dead Zone in Gulf of Mexico

Fertilizers have played a key role in the increase of food production but they key culprits in the production of greenhouse gases emissions creating dead zones

Study of fertilizers and design of Experiments

Collaborator Harold van Es

To Put It Into Perspective with Rounded Numbers….

- Average annual N_2O losses per acre (corn lands)- 7.5 lbs

  Equivalent to

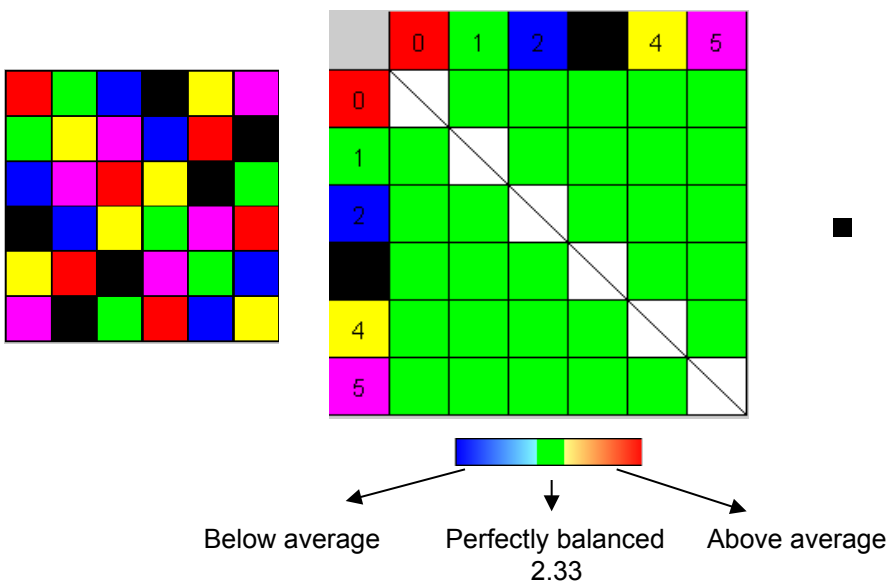
  – combustion of 126 gallons of gasoline

  – 3,700 miles of driving an average passenger car

- Assuming a farm with 500 ac of corn, the annual global warming impact is equivalent to about 1.8 million miles of driving (70 times around earth).

Design of Scientific Experiments (4 Treatments: A,B,C,D)

Latin Square



Spatially Balanced Latin Square

- Hybrid IP/CSP based
  - Assignment formulation
  - Packing formulation
  - Different CSP models
- SAT/ CSP based approach
  - State of the art model for Latin Square

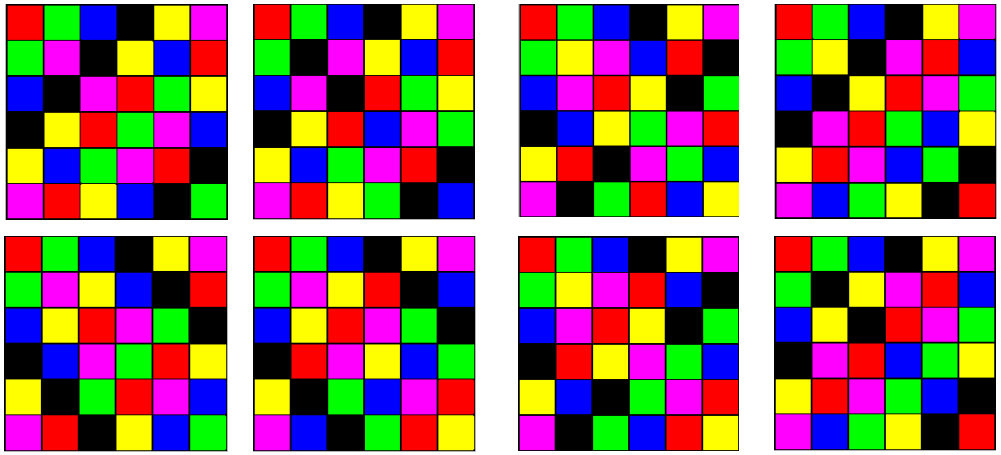
  + symmetry breaking by initializing first row and column (SBDD doesn't help; this is not a completion problem)
- Local search based approach

These approaches do not scale up

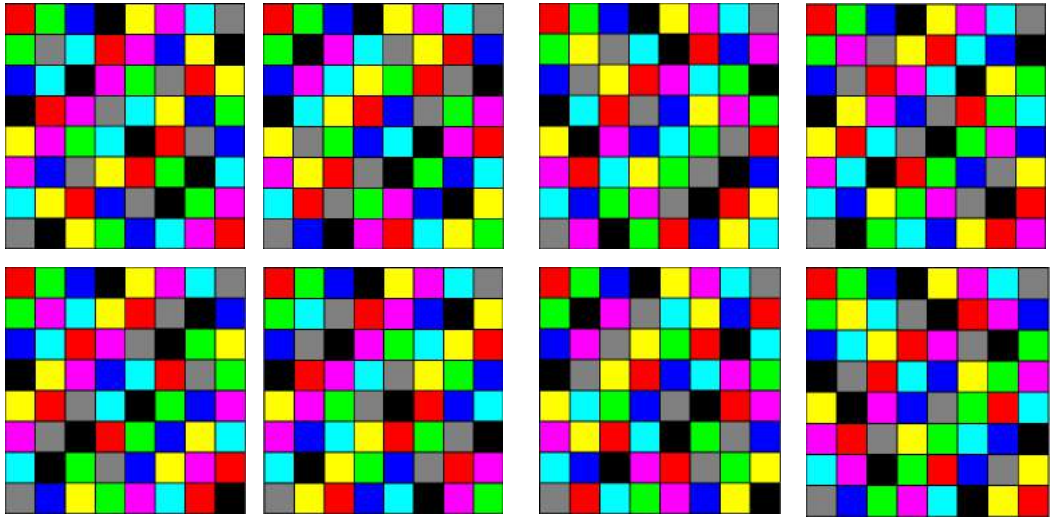
max order 6

(Target number 30).

Diagonally Symmetric SBLS



Observed symmetric SBLS of Order 6



Imposing symmetry

could generate
SBLS of order 8 and 9

Streamlining Constraint Reasoning

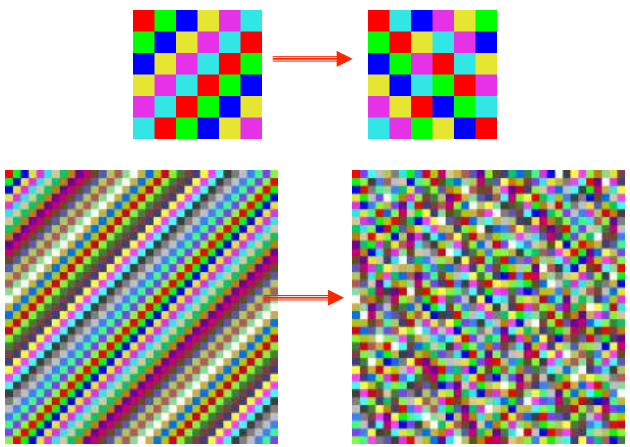
→Streamlining
→Scaling up of solutions

Discovery of structural properties across solutions .

Divide ("streamline") the search space by imposing such additional properties.

Design of Scientific Experiments for Studying Fertilizers – Spatially Balanced Latin Squares
Existence of SBLS – open problem in combinatorics

Streamlining by Permutation of Complete Columns of a Cyclic Latin Square



SBLS Order 35

Gomes , Sellmann, van Es 2003/2004

Can we find an domain independent way of streamlining?

YES: XOR-Streamlining based on random parity constraints
(Valiant and Vazirani 1986, Unique SAT)

Counting, Sampling, Inference.

Gomes, Sabharwal, Selman, 2006
Ermon,Gomes, Sabharwal, Selman, 2013, 2104, 2015

Discovery of Construction for Spatially Balanced Latin Squares: Streamlining Reasoning + Human Computation

1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	4	6	8	10	12	14	13	11	9	7	5	3	1
3	6	9	12	14	11	8	5	2	1	4	7	10	13
4	8	12	13	9	5	1	3	7	11	14	10	6	2
5	10	14	9	4	1	6	11	13	8	3	2	7	12
6	12	11	5	1	7	13	10	4	2	8	14	9	3
7	14	8	1	6	13	9	2	5	12	10	3	4	11
8	13	5	3	11	10	2	6	14	7	1	9	12	4
9	11	2	7	13	4	5	14	6	3	12	8	1	10
10	9	1	11	8	2	12	7	3	13	6	4	14	5
11	7	4	14	3	8	10	1	12	6	5	13	2	9
12	5	7	10	2	14	3	9	8	4	13	1	11	6
13	3	10	6	7	9	4	12	1	14	2	11	5	8
14	1	13	2	12	3	11	4	10	5	9	6	8	7

a	a+i

a	a-i

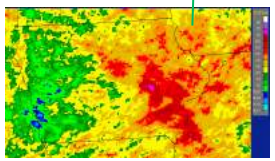
```
for row i = 1, ..., N do
    k = 1;                                  // Sequence number
    j = 1;                                  // Column index
    a_{i,j} = i;                            // First symbol of row i
    while j < N do
        if k is odd then                    // Odd sequence
            while a_{i,j} + i ≤ N and j < N do
                a_{i,j+1} = a_{i,j} + i;
                j = j + 1;
        else                                // Even sequence
            while a_{i,j} − i ≥ 1 and j < N do
                a_{i,j+1} = a_{i,j} − i;
                j = j + 1;
        if j < N then                       // Switch sequence
            if k is odd then
                a_{i,j+1} = 2N + 1 − i − a_{i,j};
            else
                a_{i,j+1} = i − a_{i,j};
            k = k + 1;
            j = j + 1;
```

Algorithm : `SBLS-sequence` procedure for SBLS of order N, when $2N + 1$ is prime.

Le Bras, Gomes, Selman AAAI 2012

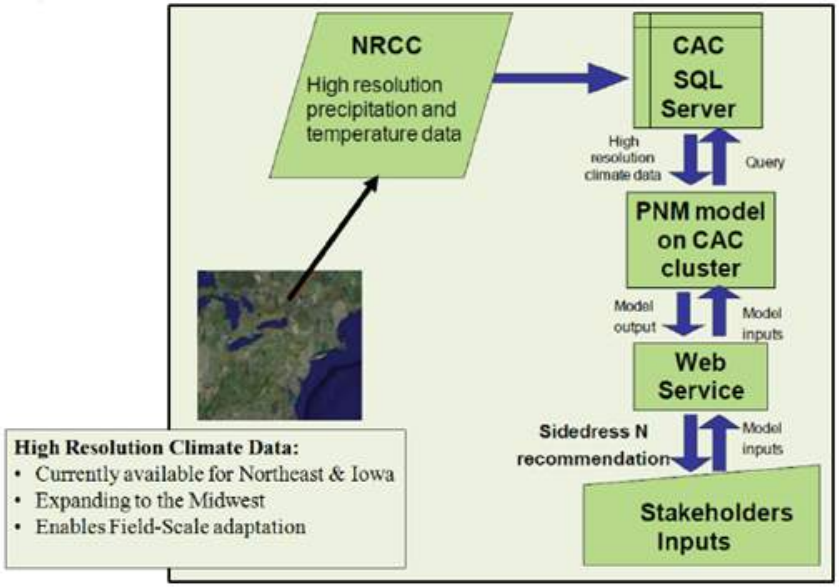
Tools for Site-Specific Adaptive Nitrogen Management

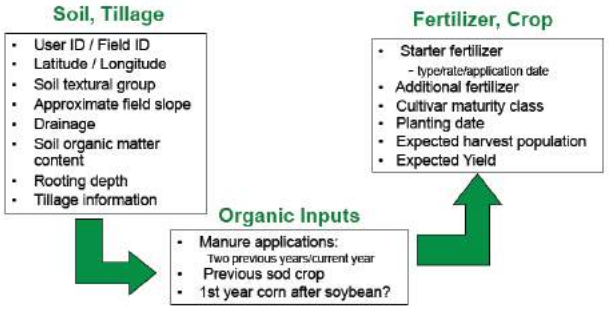
Two-Way Street: Smartphones & Tablets and the Cloud



Metereologic data

Adapt N - Infrastructure





Harold van Es

CompSustNet:
Cross-cutting Computational Themes

CompSustNet: 3 Core Sustainability Themes

Over 20 Novel Interdisciplinary Research Projects (IRPs)

I Conservation and Biodiversity

Prioritizing Conservation Decisions
Protecting Endangered Species

National Wildlife Refuge System and WaterBirds in California

Ecuador's Mega Biodiversity

Elephants and Sea Stars

II Balancing Socio-economic Needs

Protecting Farmers in Africa by Improving Weather System
Protecting Migratory Herders in Africa

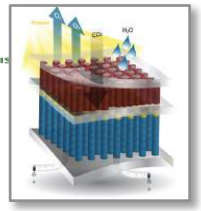
ILRI INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE
Earth Networks
TAHMO

Farmers and Migratory Herders in Africa

III Renewable Energy and sustainable materials

Accelerating the Discovery of New Sustainable Energy Materials
Energy Storage and Integration of Renewables In Smart Grid

JCAP JOINT CENTER FOR ARTIFICIAL PHOTOSYNTHESIS
Caltech
NIST
CHESS
Cornell High Energy Synchrotron

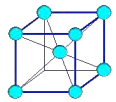


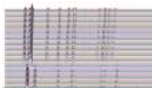
Solar Fuels

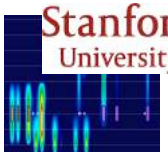
Smart Grid

Wide range of sustainability applications:
To build a ***coherent and cohesive computer science sub-field*** it's paramount to **go deep** in terms of **Cross-Cutting Core Computational Problems** and apply them to different sustainability areas.

Cross-Cutting Computational Theme:
Pattern Decomposition in Big Data







Stanford University Caltech

NIST

JCAP JOINT CENTER FOR ARTIFICIAL PHOTOSYNTHESIS

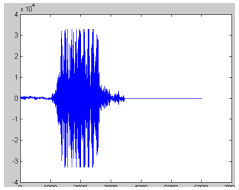
Caltech

CHESS

High Throughput Materials Structure Identification

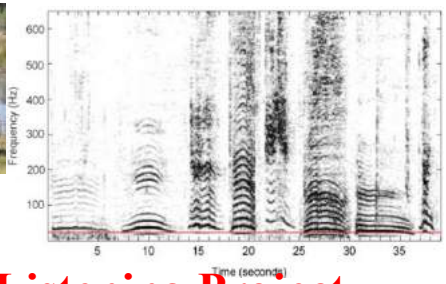
Dimensionality reduction with complex constraints

Caltech Stanford University Carnegie Mellon University



Flight Calls

Similar computational problems:
Both involve the decomposition of a signal (X-rays / audio) into patterns (crystal structures / types of sounds)



OSU Oregon State University

Elephant Listening Project

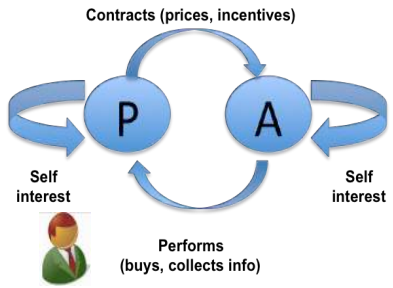
Project to monitor elephant populations by analyzing forest recordings

Our Philosophy:
Students should work on similar computational problems in different domains.
Grad students working on materials discovery also work on the elephant calls.

We represent these cross-cutting computational themes with colored "subway lines."
Pattern Decomposition in Big Data is the blue line.

Cross-Cutting Computational Theme

Findings can be transferred across domains!

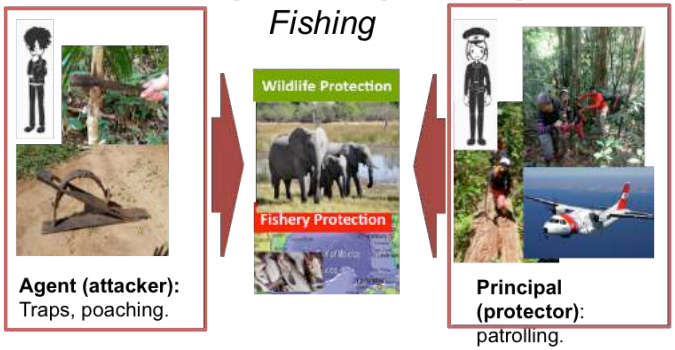


eBird Avicaching

How to incentivize Citizens to visit under-sampled areas?

Mechanism Design: Principal-Agent Game-Theory Model

Green Security Games: *Preventing Poaching and Illegal Fishing*



Where to place patrols to prevent poaching and illegal fishing?

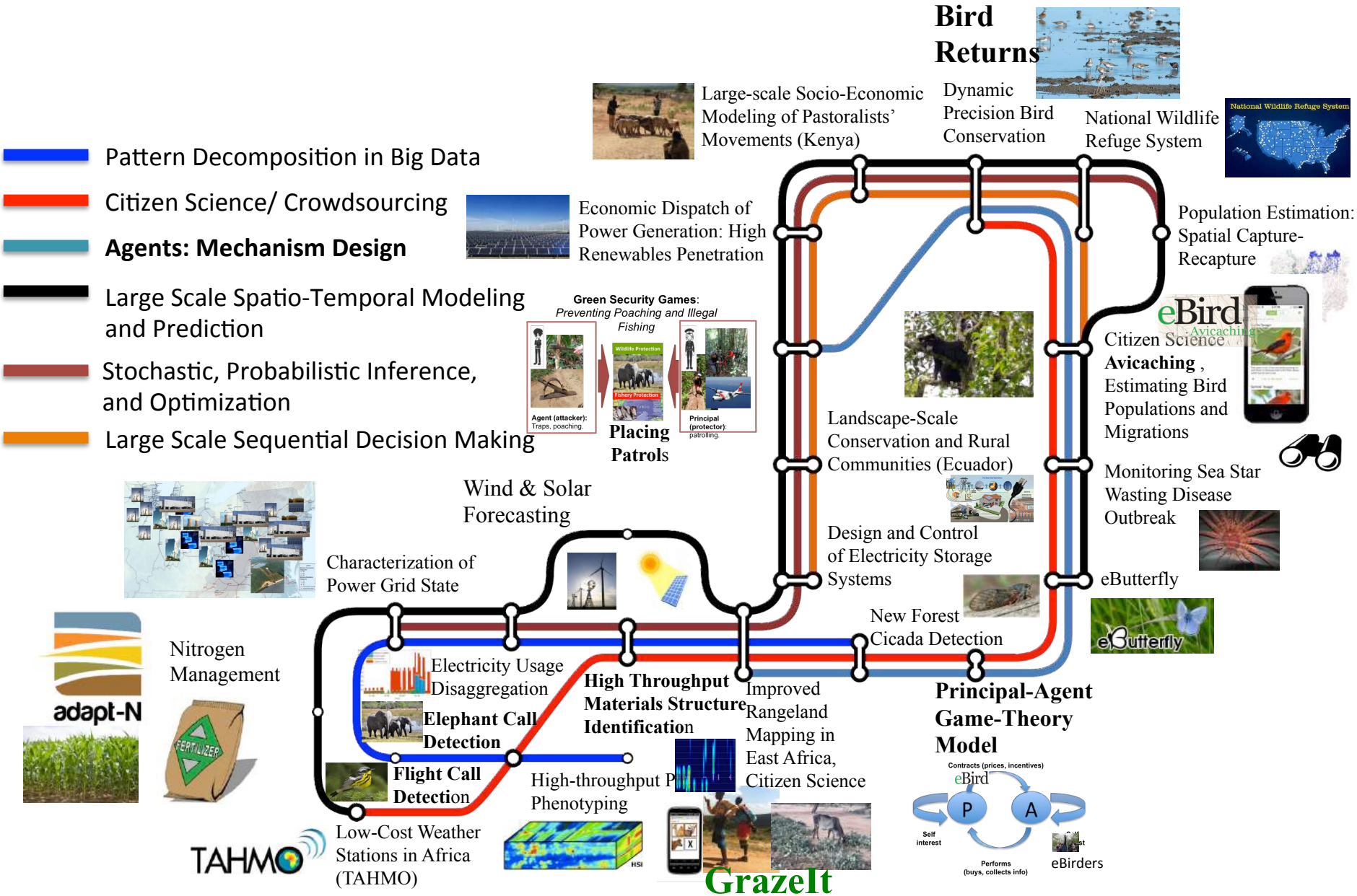
WWF

USC University of Southern California

Computationally, these problems have similarities.

We represent these cross-cutting computational themes with colored "subway lines."
Mechanism design is the baby blue line.

Subway Lines: Examples of Cross-Cutting Computational themes and Interactions of some of CompSustNet's projects



Conclusions

1st NSF Expeditions in Computational Sustainability

150+

Publications

PNAS, Science, Nature, AAAI,IJCAI, ICML, KDD, NIPS,UAI,STOC

Conference,
Referred CompSust Tracks,
Workshops

Mark Crowley
Waterloo, Canada

Theo Damoulas
Warwick, UK

Bistra Dilkina
GaTech

Stefano Ermon
Stanford

Rebecca Hutchinson
OSU

Daniel Sheldon
UMass

Gwen Spencer
Smith College

Tenure Track Assistant Professors in CS

Expedition allowed us:

- **To nucleate** the Computational Sustainability field
- **To identify** a number of **core research directions** for **maximal impact**, both in terms of **Computer Science** and **Sustainability.**

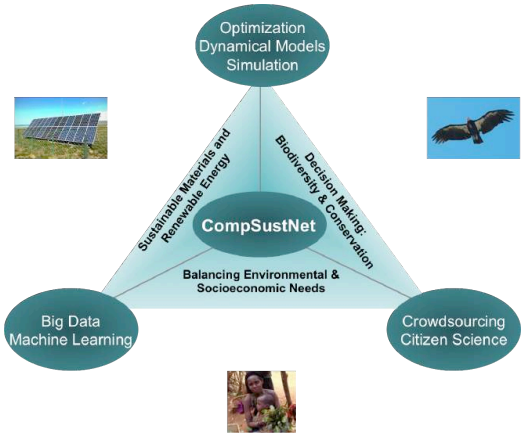
Big Challenge:

How to get involved and navigate in this highly interdisciplinary area.

Growing number of computer scientists eager to get involved in computational sustainability research, but don't have the connections and access to sustainability projects.

CompSustNet: Large Research Network for Expanding Horizons of Computational sustainbility

Research



Coordinating transformative synthesis collaborations

Interdisciplinary Research Projects (IRPs)

Research Community Building

Web Portal

Catalog Of Problems

Panels & tutorials at major conferences

Annual Conference

Host visiting Scientists

CompSust.Net

Education

Postdocs

Doctoral students

Honors projects

Research seminar series

Summer REU program targeting minority students

Distributed Course

Students' conference

Outreach

Citizen Science Projects

Engaging Gov., NGOs Institutions and Companies

K-12 Outreach

General Public Outreach

Diversity

Broad Dissemination of scientific results

Community Building

Annual CompSust Conference

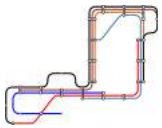
Bringing together students, researchers and practitioners

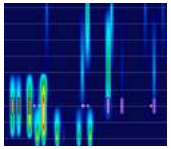
Doctoral Program and Tutorials



CompSustNet Web Portal

- Catalog of core research problems
- Datasets and open source software
- Publications
- Wiki for collaboration
- Course resources
- Discussion & e-mail list





UDiscoverIt

- Citizen science participation
- Crowdsourcing tasks



At major conferences:

Refereed CompSust Tracks, DataSet papers, Workshops, Doctoral programs, Panels and tutorials;

CompSustNet

http://compsust.net

Join computational-sustainability@yahoogroups.com

Computational Sustainability: Big Picture

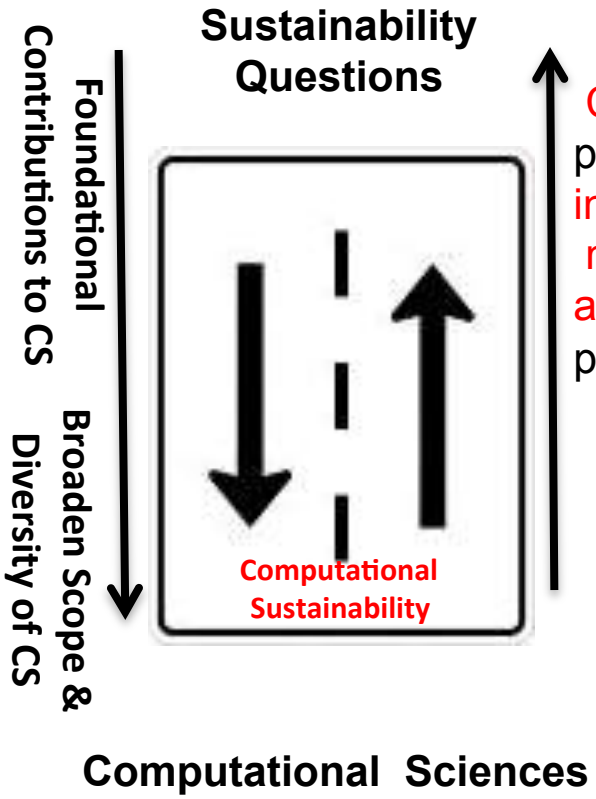
Computational Sustainability:
High potential to advance state-of-the art of Computer Science
Driven by concrete sustainability challenges

1. New challenging problems

2. New formalisms and concepts from other disciplines

→ New Core Paradigmatic problems in Comp. Sci.

→Broadening the Scope & Diversity in CS



Computational Thinking that will provide new insights, methodologies, and solutions to sustainability problems

→ Societal Impact

Thank you!