

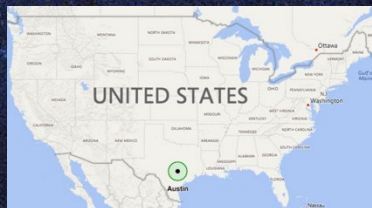
SEEKING SYNERGIES IN DATA SCIENCE AND APPLICATIONS:

How intelligent systems are driving transitions in management
for Earth Resource systems

Suzanne A. Pierce
Texas Advanced Computing Center
The University of Texas at Austin



TACC AT A GLANCE



Personnel

130+ Full time staff (~70 PhD)

General Facts

Research Division of UT Austin.

Conference & meeting capability

Key provider to XSEDE national CI

Usage Statistics

Over a Billion compute hours per year

5 Billion files, 50 Petabytes of Data,

100s of Public Datasets, 3000+ Projects

Capacity & Services

HPC, HTC, Visualization, Large scale data storage, Cloud computing

Consulting, Curation and analysis,

Code optimization, Portals and

Gateways, Web service APIs, Training

& Outreach



TACC OPERATES A professionally-supported national science cloud

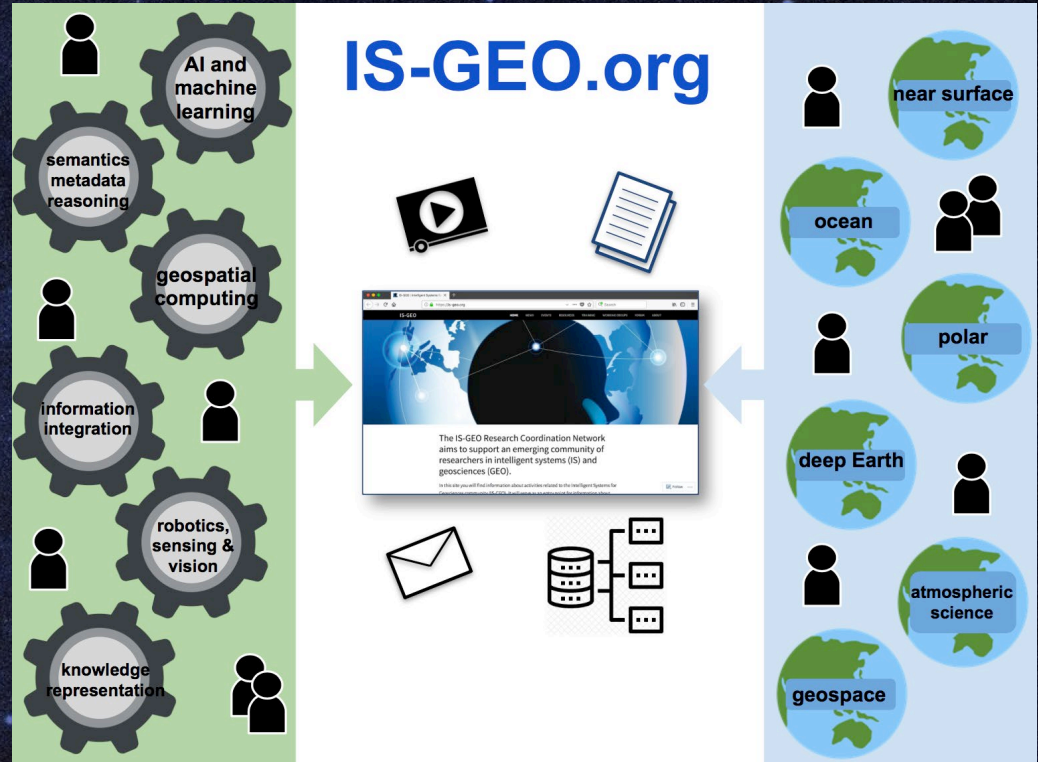


HPC, HTC, Visualization, Large scale data storage, Cloud computing, Experimental architectures



IS-GEO Research Coordination Network

Intelligent Systems and Geosciences Research Coordination Network aims to support an emerging community of researchers in intelligent systems (IS) and Geosciences



INFORMATION ABOUT EARTH IS...

COMPLEX

BIG

UNCERTAIN

CHALLENGES BECOME OPPORTUNITIES

BROAD

ILL-

STRUCTURED

DYNAMIC



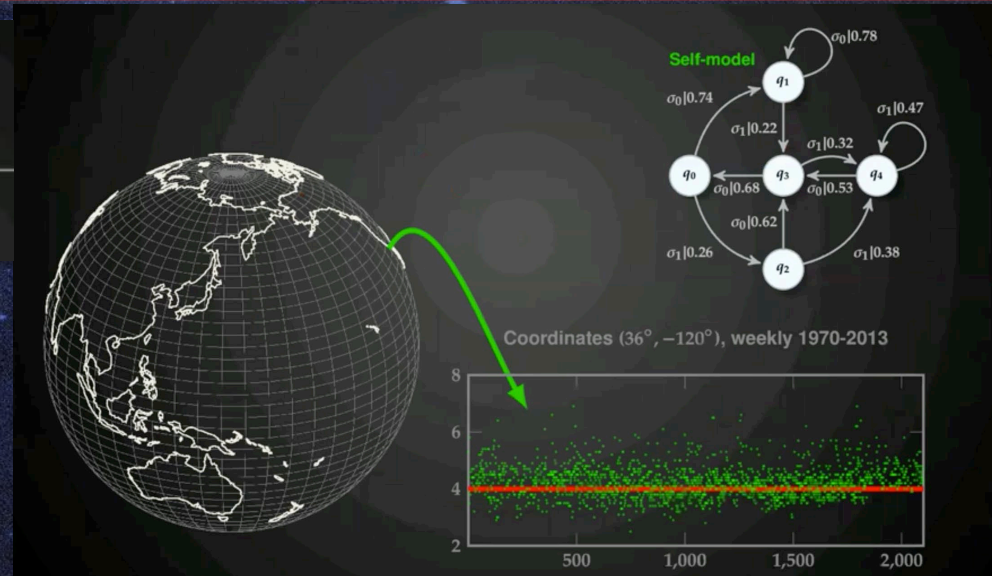
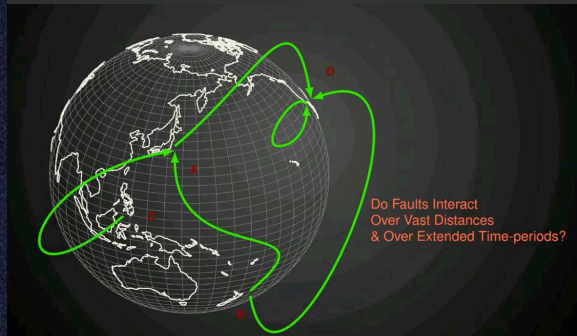
INFORMATION ABOUT EARTH IS...

BIG

PREDICTIVE PATTERN IDENTIFICATION | SPIN NETS *Machine Inference*
ISHANU CHATTOPADHYAY, University of Chicago

Predicting Seismic Events

With Both Space & Time Quantization



Zero Knowledge for Modeling Stochastic Processes



INFORMATION ABOUT EARTH IS...

BROAD

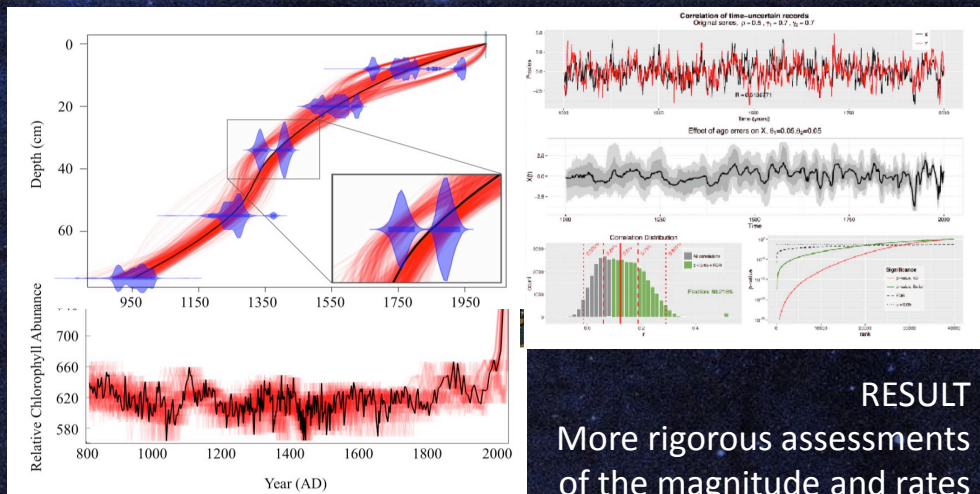
BIG

LinkedEarth is manifesting a better future for paleoscience

Data-analytic online platform that:

- (1) enables curation of publicly-accessible database by paleoclimate experts themselves,
- (2) fosters development of community standards.

HETEROGENEOUS DATA | The Future of Old Things
JULIAN EMILE-GEAY, ISI University of Southern California



Interoperable data

RESULT
More rigorous assessments
of the magnitude and rates
of pre-industrial climate
change.

MAGNITUDE & RATES OF CHANGE



INFORMATION IN THE WILD...

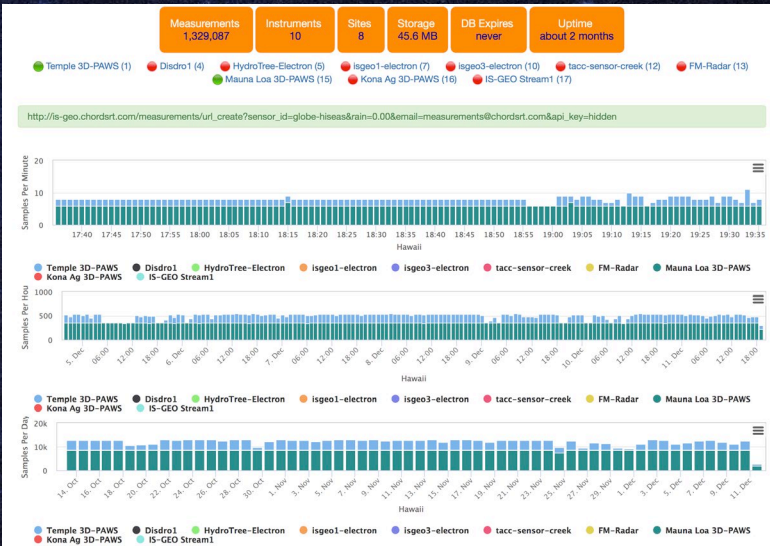
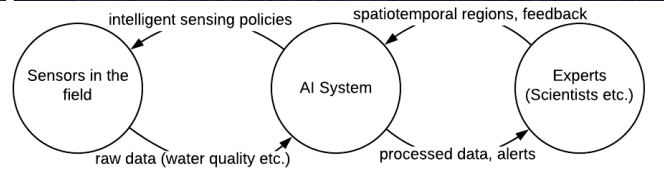
DYNAMIC

VELOCITY & SCIENCE GATEWAYS | FOSS HARDWAREX DESIGN
PETE MARCHETTO, UNIVERSITY OF MINNESOTA

BROAD
BIG



Designed & deployed 8 observation sites in ~48 hours



Data <http://is-geo.chordsrt.com/>

ADAPTIVE SENSING

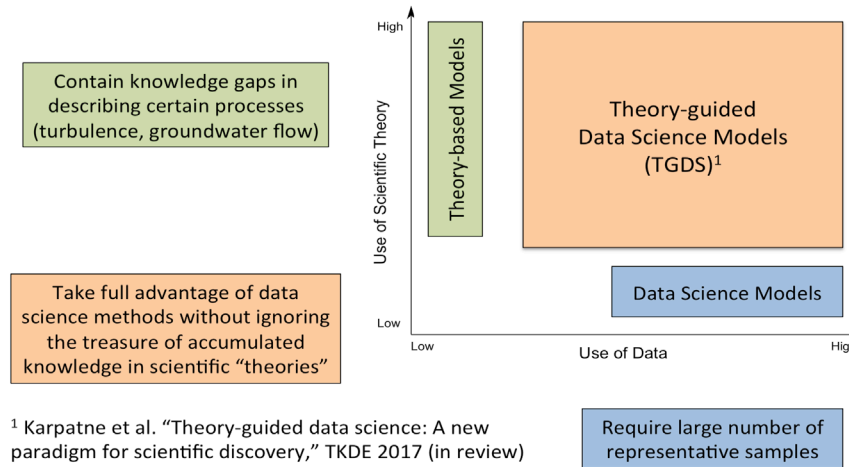


INFORMATION ABOUT EARTH IS...

ILL-
STRUCTURED
DYNAMIC
BROAD
BIG

THEORY-GUIDED | KARPATNE & KUMAR, VIRGINIA TECH AND UMN

Approaches for Scientific Discovery

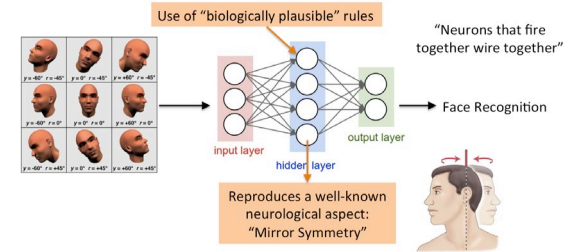


¹ Karpatne et al. "Theory-guided data science: A new paradigm for scientific discovery," TKDE 2017 (in review)

"Black-box" data science models not sufficient for knowledge discovery in scientific domains

Understanding how the human brain learns view-invariance

Leibo et al. 2016



"Black-box" data science models are insufficient



INFORMATION ABOUT EARTH IS...

UNCERTAIN

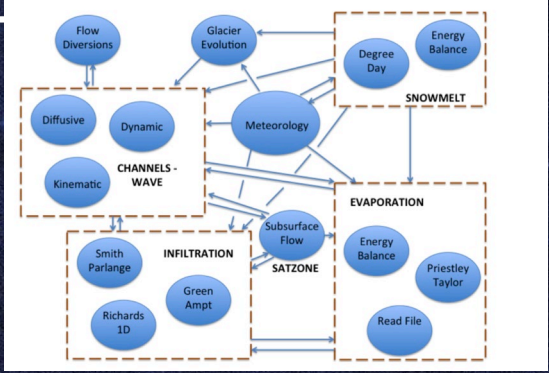
MACHINE READABILITY | Simplifying the Reuse of Datasets & Models
PECKHAM, NCAR UNIVERSITY OF COLORADO

ILL-STRUCTURED

DYNAMIC

BROAD

BIG



GSN Ontology and MCM App

Upper Ontology

CSN concept definitions as types/classes, with their predicates/properties (RDFS?)
(e.g. Root Object, Root Quantity, Object, Quantity, Operation, Variable, Process and Assumption)

Lower Ontology

Object Names List	Variable Names List
Quantity Names List	Process Names List
Operation Names List	Assumption Names List
Entity Relationships	SKOS Crosswalks
Object Q, P and A	CF Standard Names
Quantity Op, P and A	USGS P-code Names

Model Metadata from MCM App

Model 1 Metadata	Model 2 Metadata	Model 3 Metadata
------------------	------------------	------------------

Each has model-specific choices & assumptions, but the ontology is model & data agnostic.

Holding Tank for Newly Proposed Names, Name Associations & Changes

Changes are vetted. Additions integrated continuously. Other changes wait for next release.

Every blue box is a separate RDF file with assertions as S-P-O triples and may import others (TTL).

Establishing shared data models and naming



INFORMATION ABOUT EARTH IS...

COMPLEX

UNCERTAIN

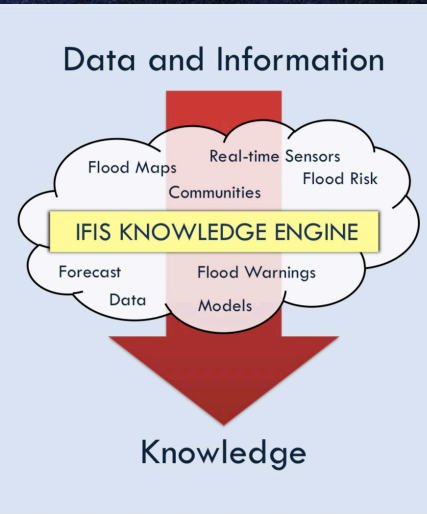
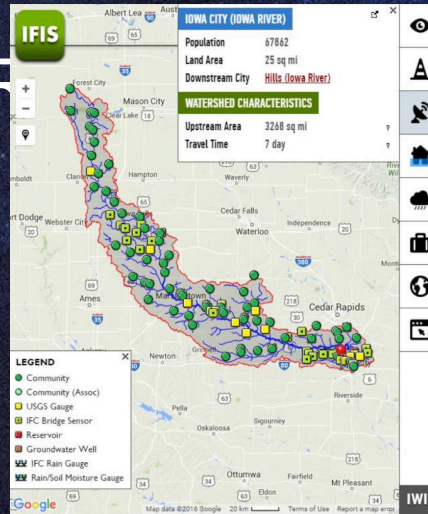
ILL-STRUCTURED

DYNAMIC

BROAD

BIG

MACHINE READABILITY | FLOOD AI ALPHA
IBRAHIM DEMIR, UNIVERSITY OF IOWA



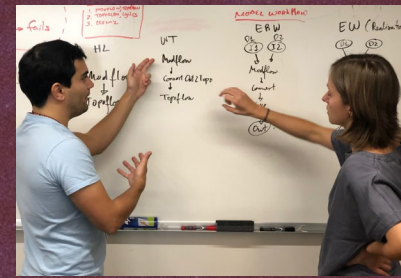
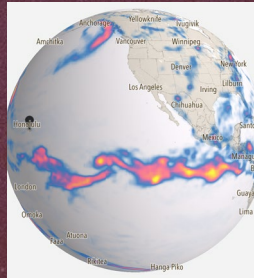
Artificial-intelligence to help the general public prepare for flooding and other natural disasters. Part of Iowa Flood Information System (IFIS), Users can ask it questions about rainfall and river levels and receive immediate answers.

AI to aid emergency response



ACCELERATING BY COLLABORATING...

- ~42 - 57 diverse participants
- 1 active volcano & hurricane Hector



Better interdisciplinary collaboration IS-GEO

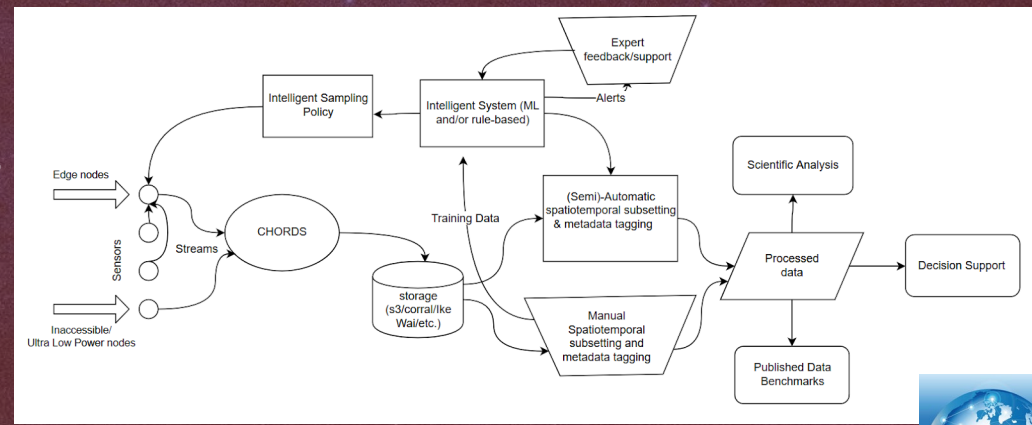
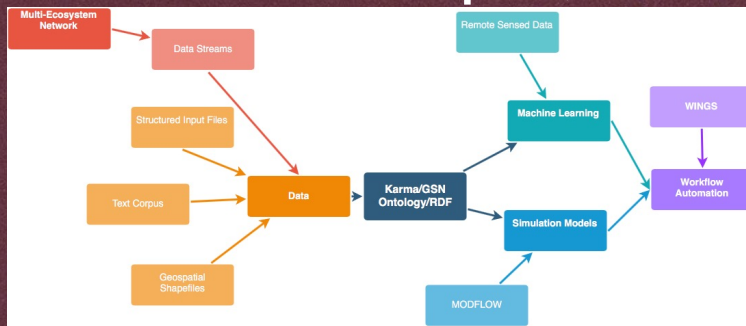


ACCELERATING SCIENCE BY SHARING...

Discovery using FM radio device

Connected new modules to IM

2-4 manuscripts, 2-3



An aerial photograph of a river with white water rapids, showing turbulent, frothy water cascading over rocks. The background is a mix of white foam and dark, rocky terrain.

PLANET TEXAS 2050

A UT Austin Grand Challenge

Cyberinfrastructure Ecosystem

To support Socio-technical and Knowledge Centric Approaches for Complex, Resilience Problems

A satellite-style map of Texas is the background. A semi-transparent grey box is overlaid on the top left, containing the title and subtitle. The text is in white and orange. The map shows the state's outline and major geographical features like rivers and the coastline.

PLANET TEXAS 2050

A UT Austin Grand Challenge

Texas' population is projected to nearly double by 2050.

We will experience more extreme weather events: more floods, more droughts, and more heat.

We need to better manage our state's resources to support these demands.

Making Texas resilient is our grand challenge.

Vision: a healthy, safe, just, ecologically + economically vibrant Texas by 2050

To get there

- Fundamental sciences
- Applied sciences
- Health sciences
- Behavioral sciences
- Humanities & communications
- Engineering
- Planning
- More
- Past, present, future



DataX
Integration

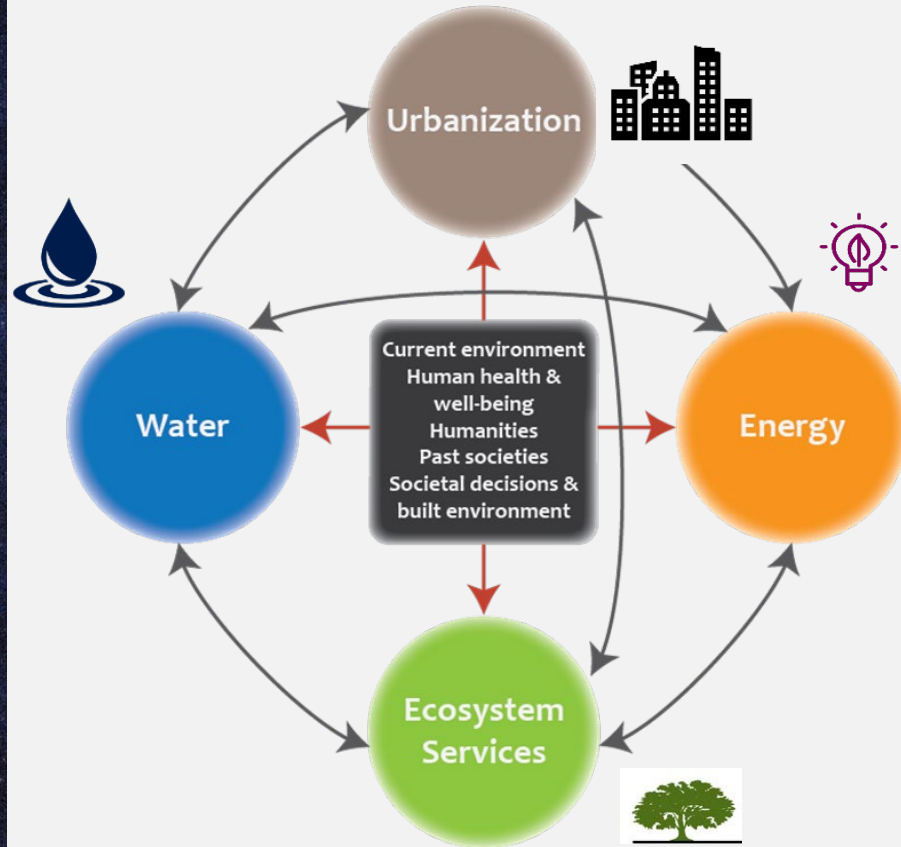
Model
Development

Simulations,
Strategies, and
Policies

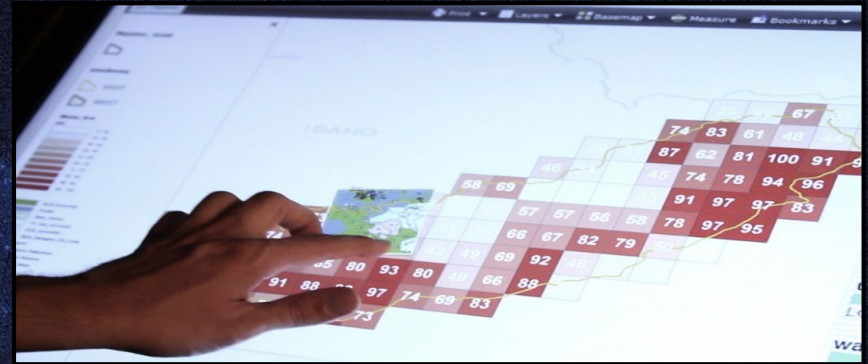
Climate Change

+

Population Growth



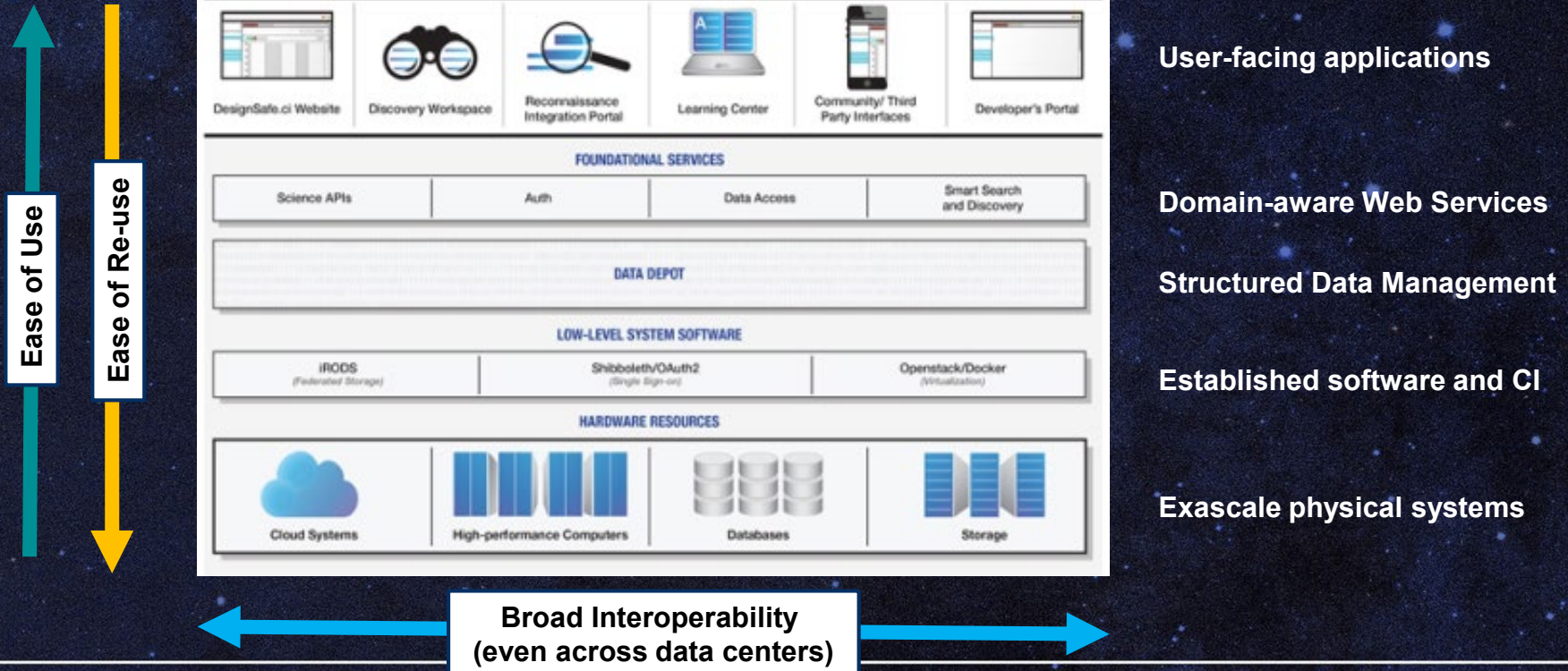
How can scientists present information to stakeholders and decision makers in a way that supports dialogue And leads to science-based decision making?



What features and specifications can create a reusable and intelligent decision support system?

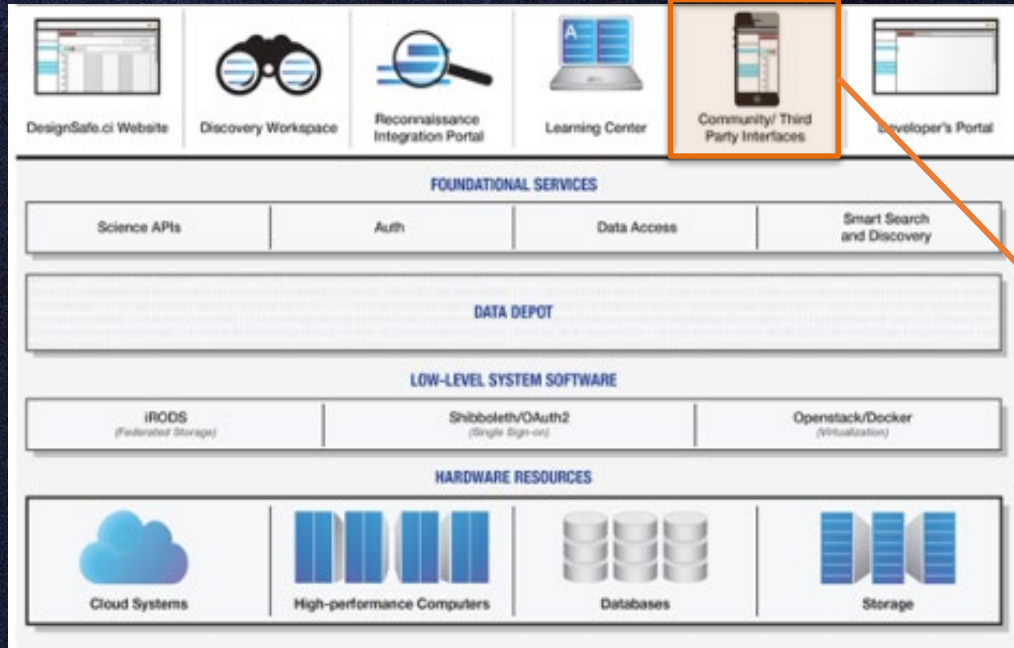
Designing intelligent Decision Support Systems

TACC OPERATES A professionally-supported national science cloud / PT2050 is leveraging it

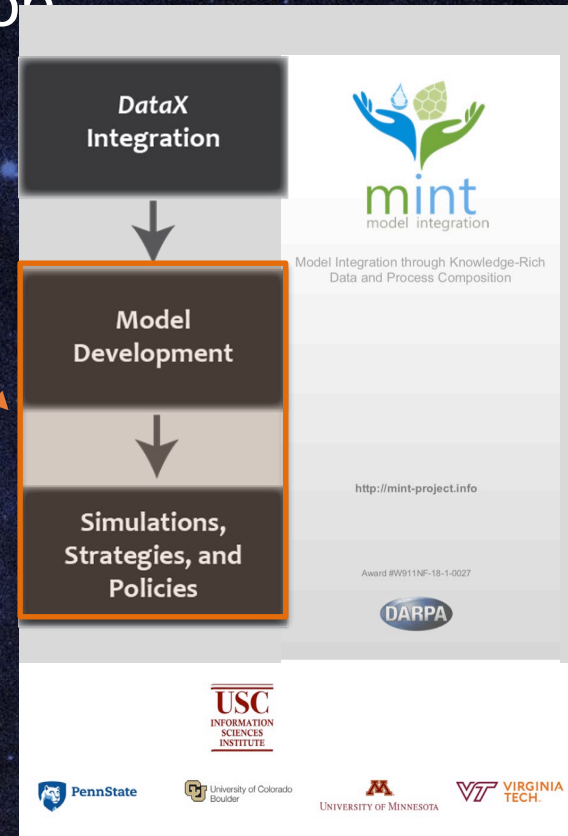


INTEGRATED MODELING with linked data and model catalogs connect to workflows & HPC job queues

Ease of Use
Ease of Re-use



Broad Interoperability (even across data centers)



Cyberinfrastructure Ecosystem

PLANET TEXAS 2050

A UT Grand Challenge



sawp33



Dashboard



Data Files



Applications



Notifications



Search



Rename



Move



Copy



Preview



Download



Move to Trash

+ Add

MY DATA / PT2050 / PT2050 RESEARCH

My Data

Community Data

Name	Size	Last Modified	Permissions
Ⓛ .	4.0 kB	10/13/18 1:49 PM	All

DataX and Integrated Modeling (IM) architecture use Data Lifecycle as an organizing concept

Cyberinfrastructure Ecosystem

PLANET TEXAS 2050

A UT Grand Challenge



sawp33



Dashboard



Data Files



Applications



Notifications



Search

- Rename
- Move
- Copy
- Preview
- Download
- Move to Trash

+ Add

MY DATA / PT2050 / PT2050 RESEARCH

- My Data
- Community Data

- Data Streams
- Integrated Modeling

Name	Size	Last Modified	Permissions
.	4.0 kB	10/13/18 1:49 PM	All

Cyberinfrastructure Ecosystem

PLANET TEXAS 2050

A UT Grand Challenge



sawp33



Dashboard



Data Files



Applications



Notifications



Search

- Rename
- Move
- Copy
- Preview
- Download
- Move to Trash

+ Add

MY DATA / PT2050 / PT2050 RESEARCH

- My Data
- Community Data

- Data Streams
- Integrated Modeling

Name	Size	Last Modified	Permissions
.	4.0 kB	10/13/18 1:49 PM	All
T10 Graybelt	4.0 kB	10/13/18 1:47 PM	All
T11 Averaging Water	32.0 kB	11/9/18 10:31 AM	All
T12 Transportation Air	4.0 kB	10/13/18 1:49 PM	All
T6 Urban Observatory	4.0 kB	10/13/18 1:45 PM	All
T7 Premodern Urban Environments	4.0 kB	10/13/18 1:46 PM	All
T8 Urban Watersheds	4.0 kB	10/13/18 1:45 PM	All
T9 Water Stories	4.0 kB	10/13/18 1:48 PM	All

Cyberinfrastructure Ecosystem

PLANET TEXAS 2050

A UT Grand Challenge


 sawp33



Dashboard



Data Files



Applications



Notifications








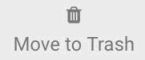
Search

+ Add

- My Data
- Community Data

- Data Streams
- Integrated Modeling

Search

MY DATA / PT2050 / PT2050 RESEARCH

Name	Size	Last Modified	Permissions
.	4.0 kB	10/13/18 1:49 PM	All
T10 Graybelt	4.0 kB	10/13/18 1:47 PM	All
T11 Averaging Water	32.0 kB	11/9/18 10:31 AM	All
T12 Transportation Air	4.0 kB	10/13/18 1:49 PM	All
T6 Urban Observatory	4.0 kB	10/13/18 1:45 PM	All
T7 Premodern Urban Environments	4.0 kB	10/13/18 1:46 PM	All
T8 Urban Watersheds	4.0 kB	10/13/18 1:45 PM	All
T9 Water Stories	4.0 kB	10/13/18 1:48 PM	All



Concept Model to CAG



Simulation [0]

Visualization [0]

Data Processing [1]

Utilities [2]

My Apps [1]



Dashboard



Data Files



Applications

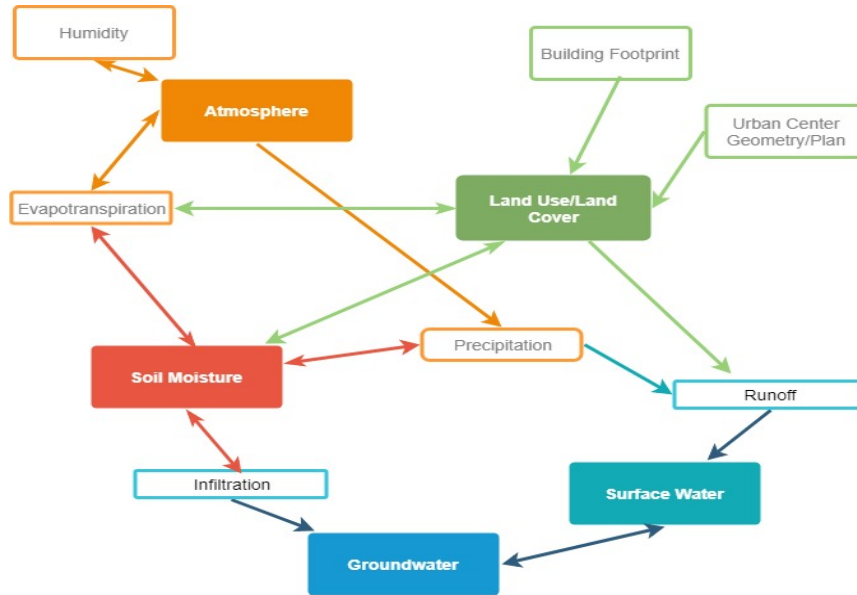


Notifications



Search

ANALYZE



High Level Workflow Template

PLANET TEXAS 2050
A UT Grand Challenge

Simulation [0] Visualization [0] Data Processing [1] Utilities [2] My Apps [1]

Dashboard

Data Files

Applications

Notifications

Search

ANALYZE

```

    graph TD
      Humidity --> Atmosphere
      Atmosphere --> Evapotranspiration
      Atmosphere --> Precipitation
      Atmosphere --> LandUse[Land Use/Land Cover]
      Evapotranspiration --> LandUse
      BuildingFootprint[Building Footprint] --> LandUse
      UrbanCenter[Urban Center Geometry/Plan] --> LandUse
      LandUse --> Precipitation
      LandUse --> Runoff
      Precipitation --> SoilMoisture
      Precipitation --> Runoff
      SoilMoisture --> Evapotranspiration
      SoilMoisture --> Infiltration
      Runoff --> SurfaceWater
      Infiltration --> Groundwater
      SurfaceWater --> Groundwater
      Groundwater --> SurfaceWater
  
```

1. PRISM2RCH – MODFLOW
2. CESM22RCH – MODFLOW
3. PRISM2RCH-NLDAS-MODFLOW
4. CESM22RCH –SoilModel-MODFLOW
5. PRISM2RCH-NLDAS-HAND-MODFLOW
6. CESM22RCH –SoilModel-NWM- MODFLOW
7. ESM22RCH –SoilModel-SWAT- MODFLOW

High Level Workflow Template

PLANET TEXAS 2050
A UT Grand Challenge

Simulation [0] Visualization [0] Data Processing [1] Utilities [2] My Apps [1]

Dashboard

Data Files

Applications

Notifications

Search

ANALYZE

```
graph TD; Humidity --> Atmosphere; Atmosphere --> Evapotranspiration; Atmosphere --> Precipitation; Atmosphere --> Soil Moisture; Evapotranspiration --> Atmosphere; Evapotranspiration --> Land Use/Land Cover; Land Use/Land Cover --> Precipitation; Land Use/Land Cover --> Runoff; Land Use/Land Cover --> Soil Moisture; Building Footprint --> Land Use/Land Cover; UrbanCenter[Urban Center Geometry/Plan] --> Land Use/Land Cover; Precipitation --> Soil Moisture; Precipitation --> Runoff; Soil Moisture --> Evapotranspiration; Soil Moisture --> Infiltration; Infiltration --> Groundwater; Runoff --> SurfaceWater[Surface Water]; SurfaceWater --> Groundwater; Groundwater --> SurfaceWater;
```

1. PRISM2RCH – MODFLOW
2. CESM22RCH – MODFLOW
3. PRISM2RCH-NLDAS-MODFLOW
4. CESM22RCH –SoilModel-MODFLOW
5. PRISM2RCH-NLDAS-HAND-MODFLOW
6. CESM22RCH –SoilModel-NWM- MODFLOW
7. ESM22RCH –SoilModel-SWAT- MODFLOW

Data Ingest → Storage → Analyze

High Level Workflow Template

PLANET TEXAS 2050
A UT Grand Challenge

Simulation [0] Visualization [0] Data Processing [1] Utilities [2] My Apps [1]

Dashboard

Data Files

Applications

Notifications

Search

ANALYZE

1. PRISM2RCH – MODFLOW
2. CESM22RCH – MODFLOW
3. PRISM2RCH-NLDAS-MODFLOW
4. CESM22RCH –SoilModel-MODFLOW
5. PRISM2RCH-NLDAS-HAND-MODFLOW
6. CESM22RCH –SoilModel-NWM- MODFLOW
7. ESM22RCH –SoilModel-SWAT- MODFLOW

High Level Workflow Template

PLANET TEXAS 2050
A UT Grand Challenge

Simulation [0] Visualization [0] Data Processing [1] Utilities [2] My Apps [1]

Dashboard

Data Files

Applications

Notifications

Search

ANALYZE

```

graph TD
    Humidity[Humidity] <--> Atmosphere[Atmosphere]
    BuildingFootprint[Building Footprint] --> LandUse[Land Use/Land Cover]
    UrbanCenter[Urban Center Geometry/Plan] --> LandUse
    LandUse <--> Atmosphere
    LandUse <--> Evapotranspiration[Evapotranspiration]
    LandUse <--> SoilMoisture[Soil Moisture]
    LandUse <--> Precipitation[Precipitation]
    LandUse <--> Runoff[Runoff]
    Atmosphere <--> Evapotranspiration
    Evapotranspiration <--> SoilMoisture
    Precipitation <--> SoilMoisture
    SoilMoisture <--> Infiltration[Infiltration]
    Infiltration <--> Groundwater[Groundwater]
    Runoff <--> SurfaceWater[Surface Water]
    SurfaceWater <--> Groundwater
  
```

1. PRISM2RCH – MODFLOW
2. CESM22RCH – MODFLOW
3. PRISM2RCH-NLDAS-MODFLOW
4. CESM22RCH –SoilModel-MODFLOW
5. PRISM2RCH-NLDAS-HAND-MODFLOW
6. CESM22RCH –SoilModel-NWM- MODFLOW
7. ESM22RCH –SoilModel-SWAT- MODFLOW



Execution Ready Workflow

PLANET TEXAS 2050

A UT Grand Challenge



sawp33



Dashboard



Data Files



Applications



Notifications



Search

Simulation [0]

Visualization [0]

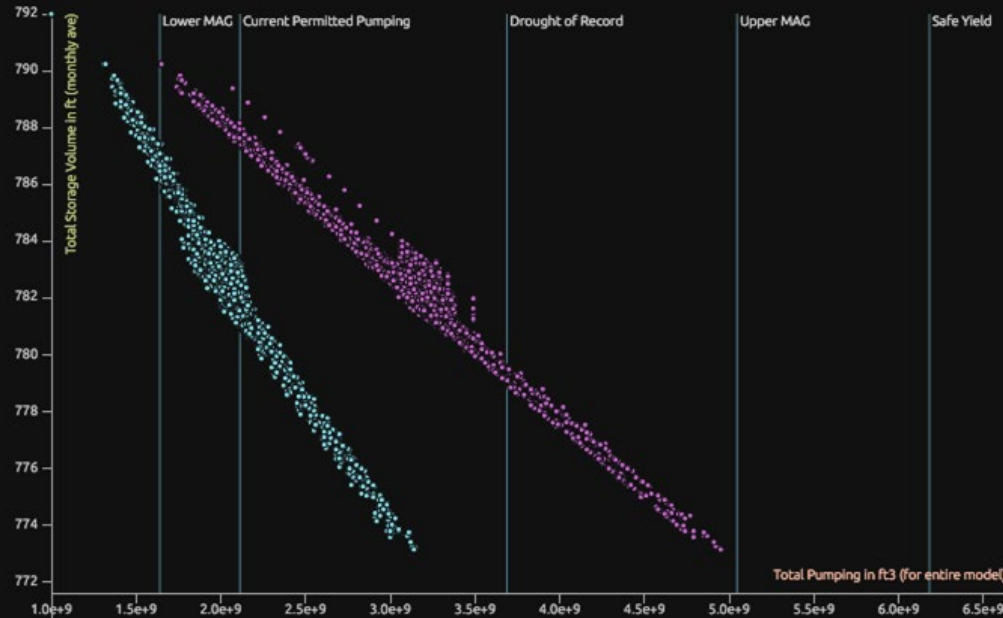
Data Processing [1]

Utilities [2]

My Apps [1]

ANALYZE

Identifying Desired Future Conditions (DFCs)



1. PRISM2RCH – MODFLOW

2. CESM22RCH – MODFLOW

3. PRISM2RCH-NLDAS-MODFLOW

4. CESM22RCH –SoilModel-MODFLOW

5. PRISM2RCH-NLDAS-HAND-MODFLOW

6. CESM22RCH –SoilModel-NWM- MODFLOW

7. ESM22RCH –SoilModel-SWAT- MODFLOW



