



**Pacific
Northwest**
NATIONAL LABORATORY

Facilitating Collaboration Across the MultiSector Dynamics Community

**Casey Burleyson, Ian Kraucunas,
Carina Lansing, and Dave Millard**

U.S. DEPARTMENT OF
ENERGY **BATTELLE**

PNNL is operated by Battelle for the U.S. Department of Energy



Challenges to Collaborative Data and Code Management

1) As projects have grown in size and complexity, there is increasingly a need for data and code to be developed in open environments and managed in ways that promote repeatability, reusability, and community adoption.



Modeling Teams



Multi-model Experiments

Projects



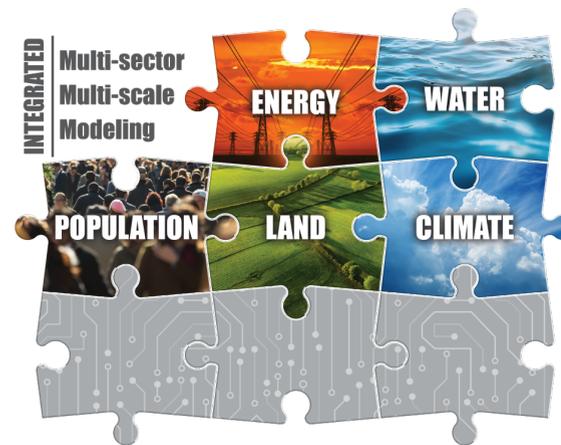
MSD Community



Challenges to Collaborative Data and Code Management

1) As projects have grown in size and complexity, there is increasingly a need for data and code to be developed in open environments and managed in ways that promote repeatability, reusability, and community adoption.

2) Collaborators increasingly span multiple institutions, each with their own computational environments, data archives, collaboration tools, and access restrictions, making it more difficult for teams to collaborate and models to communicate.



Challenges to Collaborative Data and Code Management

- 1) As projects have grown in size and complexity, there is increasingly a need for data and code to be developed in open environments and managed in ways that promote repeatability, reusability, and community adoption.
- 2) Collaborators increasingly span multiple institutions, each with their own computational environments, data archives, collaboration tools, and access restrictions, making it more difficult for teams to collaborate and models to communicate.
- 3) There is a lack of community tools for documenting, archiving, and sharing datasets, along with a culture that does not tend to promote data reuse or extensibility.

Challenges to Collaborative Data and Code Management

- 1) As projects have grown in size and complexity, there is increasingly a need for data and code to be developed in open environments and managed in ways that promote repeatability, reusability, and community adoption.
- 2) Collaborators increasingly span multiple institutions, each with their own computational environments, data archives, collaboration tools, and access restrictions, making it more difficult for teams to collaborate and models to communicate.
- 3) There is a lack of community tools for documenting, archiving, and sharing datasets, along with a culture that does not tend to promote data reuse or extensibility.
- 4) There is limited use of model and workflow versioning, which constrains the repeatability and extensibility of individual numerical experiments.

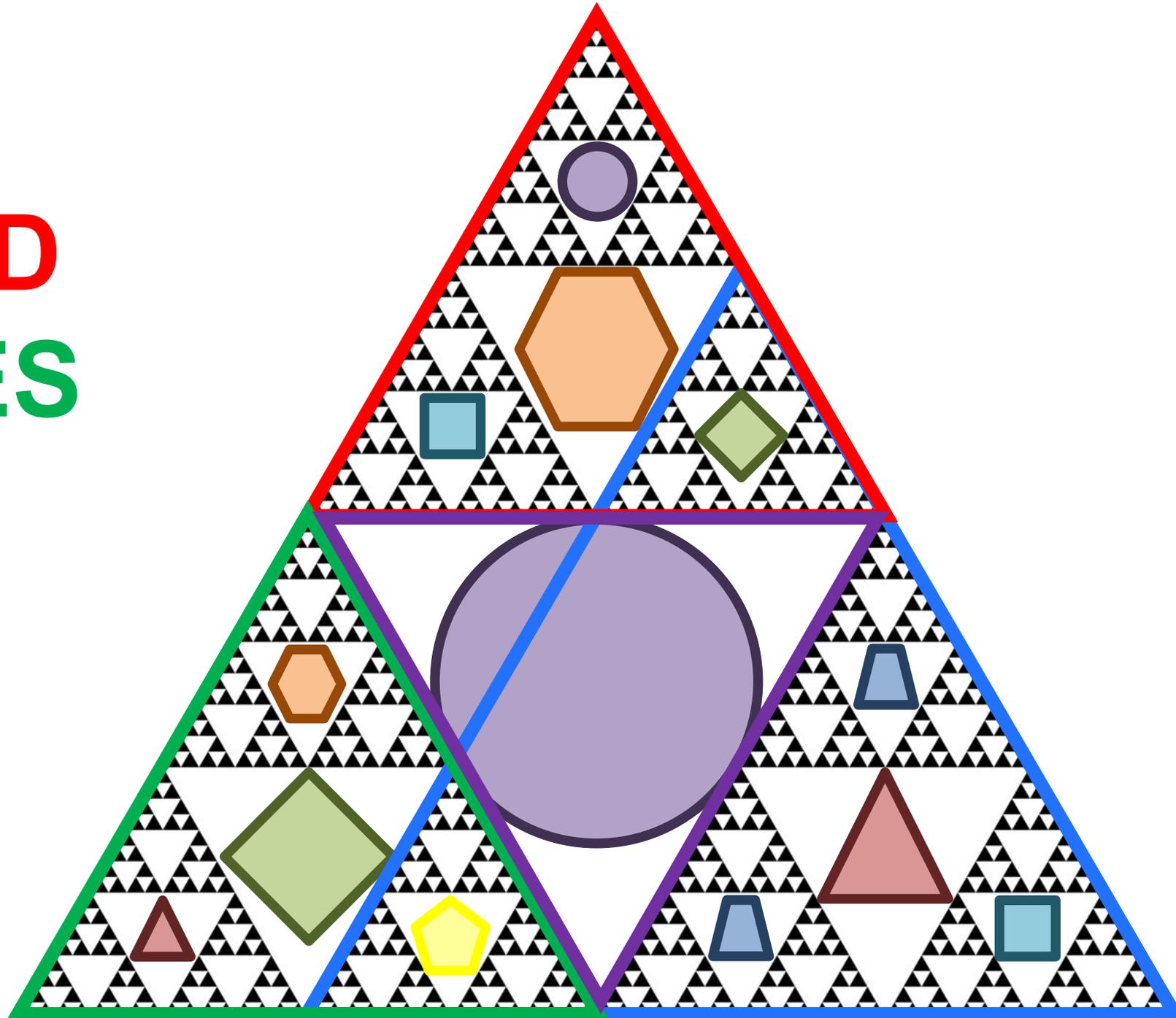
21st century United States emissions mitigation could increase water stress more than the climate change it is mitigating

Mohamad I. Hejazi^{a,1}, Nathalie Voisin^b, Lu Liu^a, Lisa M. Bramer^b, Daniel C. Fortin^b, John E. Hathaway^b, Maoyi Huang^b, Page Kyle^a, L. Ruby Leung^b, Hong-Yi Li^b, Ying Liu^b, Pralit L. Patel^a, Trenton C. Pulsipher^b, Jennie S. Rice^b, Teklu K. Tesfa^b, Chris R. Vernon^b, and Yuyu Zhou^a

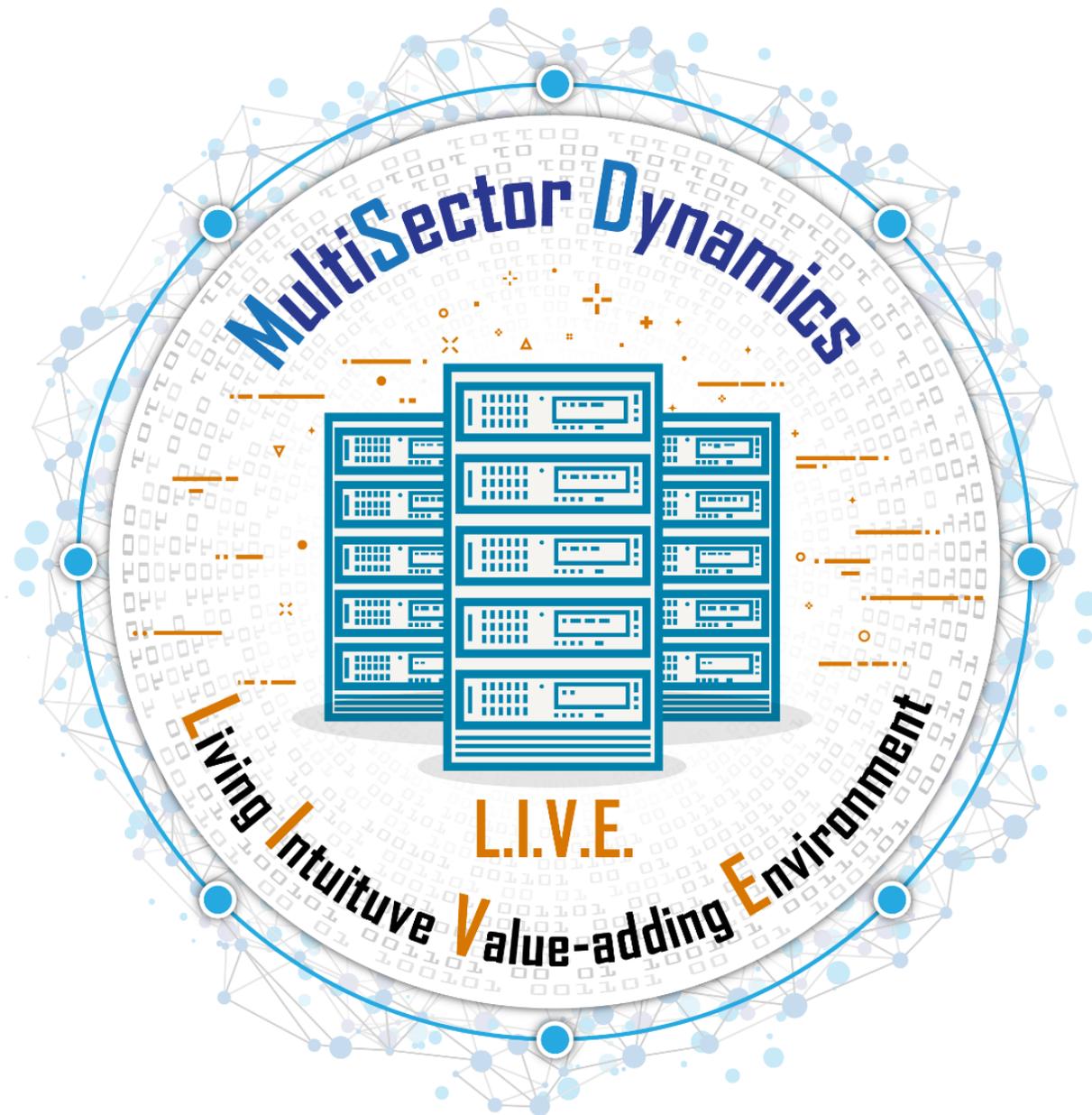
^aJoint Global Change Research Institute, Pacific Northwest National Laboratory/University of Maryland, College Park, MD 20740; and ^bPacific Northwest National Laboratory, Richland, WA 99352

Sharing Data and Code Across the MSD Community of Practice

IM3
IHESD
PCHES
MIT

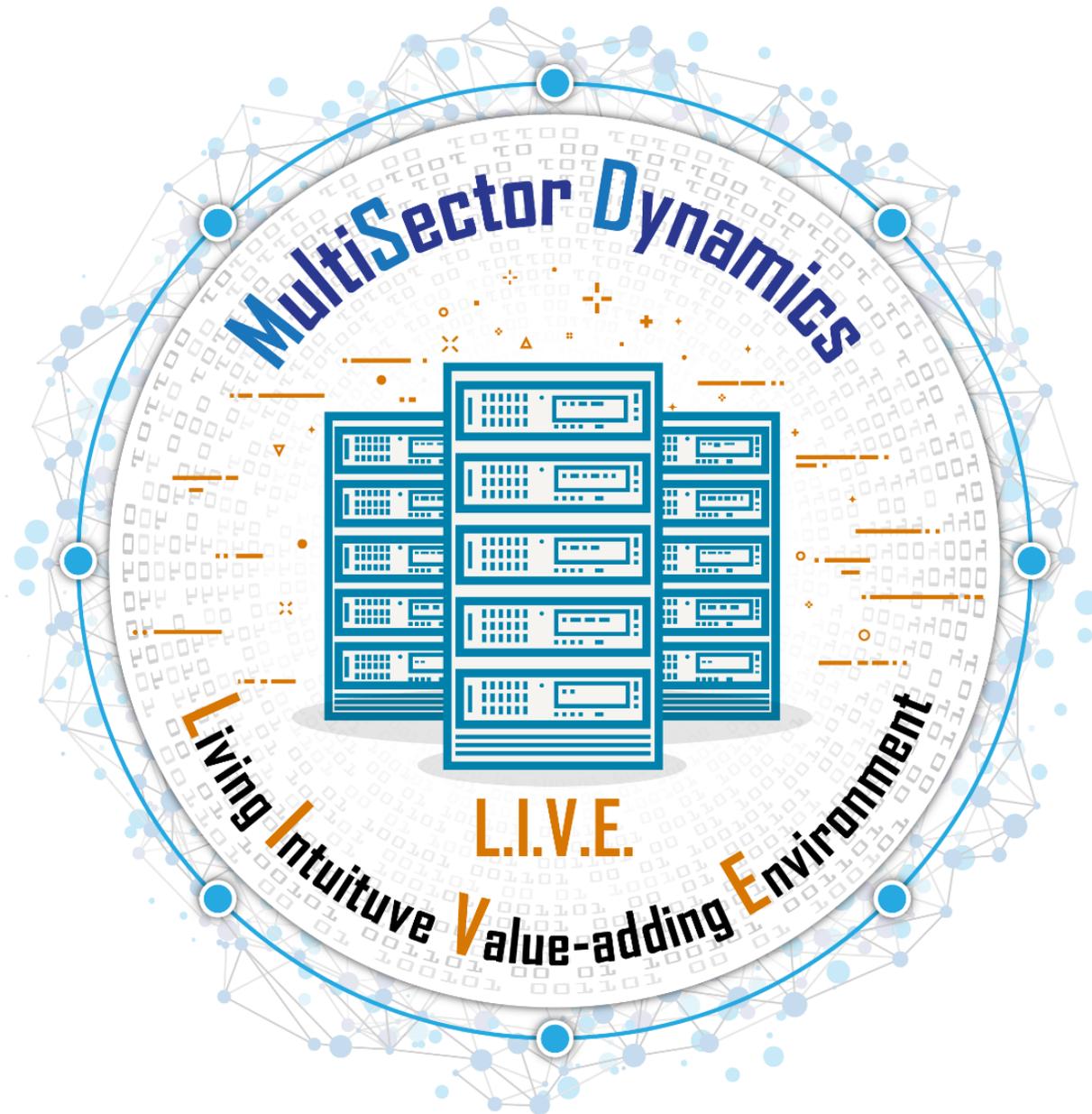


What is MSD-LIVE?



The MultiSector Dynamics – Living, Intuitive, Value-adding, Environment (MSD-LIVE) is a flexible data and code management system combined with a computational platform and set of user interfaces that will enable MSD researchers to document and archive their data, run their models and analysis tools, and share their data, software, and multi-model workflows.

What is MSD-LIVE?

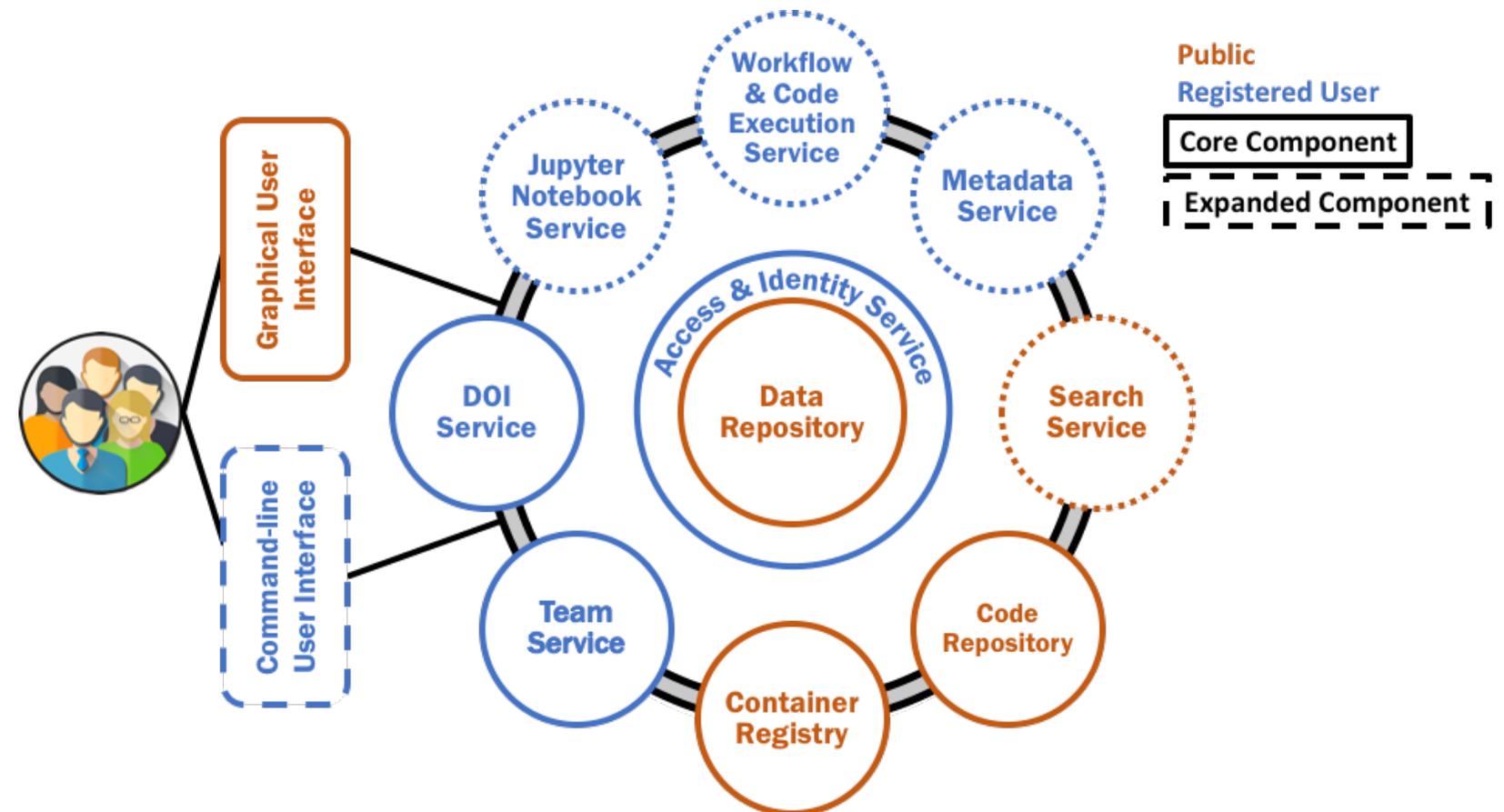


The MultiSector Dynamics – Living, Intuitive, Value-adding, Environment (MSD-LIVE) is a flexible data and code management system combined with a computational platform and set of user interfaces that will enable MSD researchers to document and archive their data, run their models and analysis tools, and share their data, software, and multi-model workflows.

What is MSD-LIVE?

The MultiSector Dynamics – Living, Intuitive, Value-adding, Environment (MSD-LIVE) is a flexible data and code management system combined with a computational platform and set of user interfaces that will enable MSD researchers to document and archive their data, run their models and analysis tools, and share their data, software, and multi-model workflows.

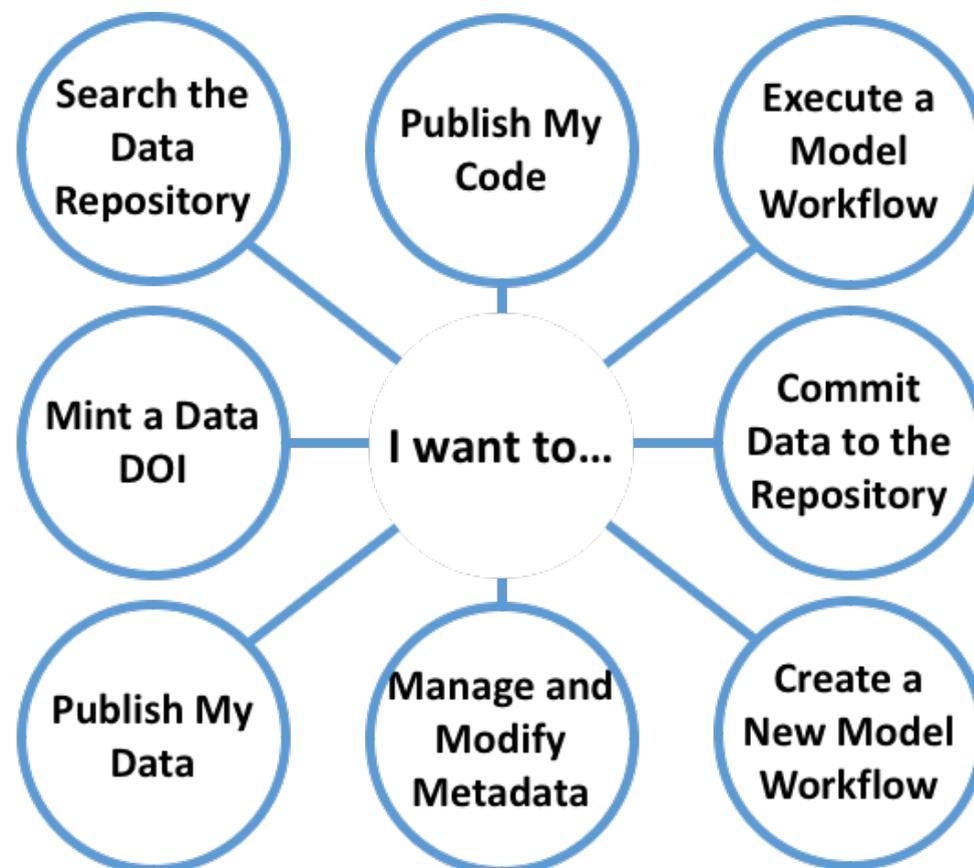
We envision the architecture as a collection of services or components accessed via two potential interfaces.



What is MSD-LIVE?

The MultiSector Dynamics – Living, Intuitive, Value-adding, Environment (MSD-LIVE) is a flexible data and code management system combined with a computational platform and set of user interfaces that will enable MSD researchers to document and archive their data, run their models and analysis tools, and share their data, software, and multi-model workflows.

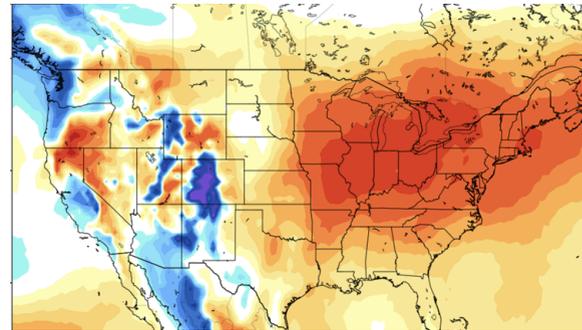
- View my:**
- I) Teams
 - II) Data
 - III) Code
 - IV) Workflows
 - V) Jupyter Notebooks
 - VI) Docker Containers



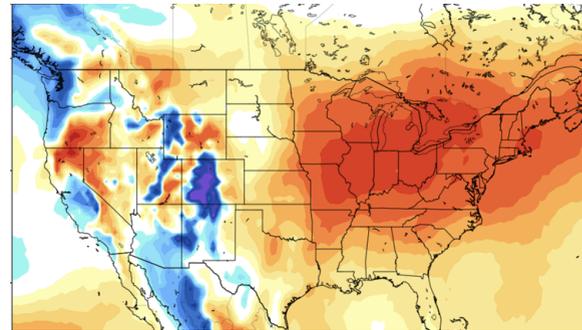
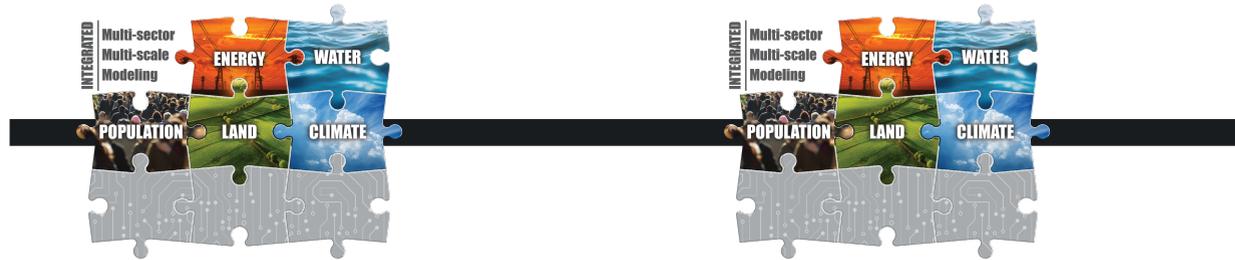
A potential task-focused landing page that a registered user would see upon logging into MSD-LIVE.



How Might MSD-LIVE Benefit You?



How Might MSD-LIVE Benefit You?



MOSART-WM

+

GCAM

+

BEND

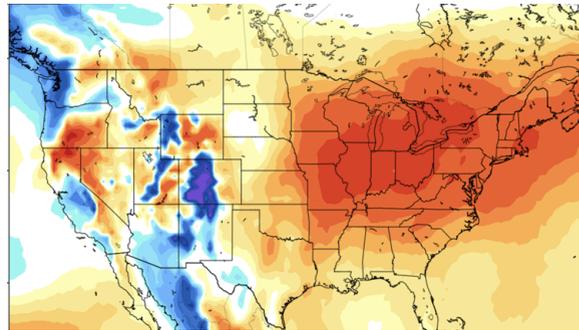
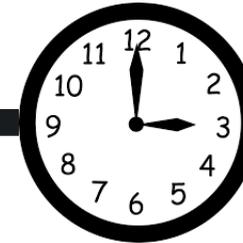
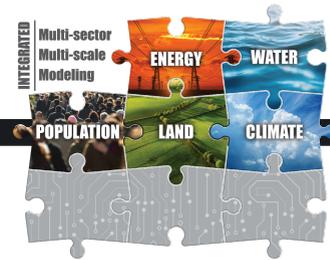
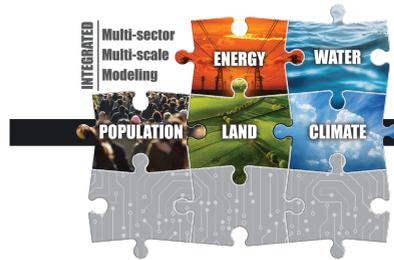
+

PLEXOS

=

nature

How Might MSD-LIVE Benefit You?



MOSART-WM

+

GCAM

+

BEND

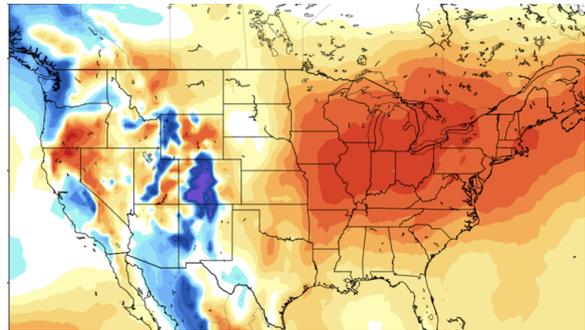
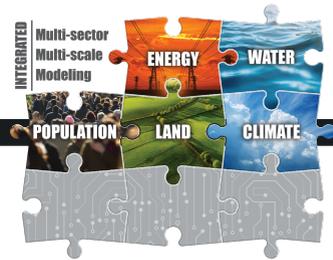
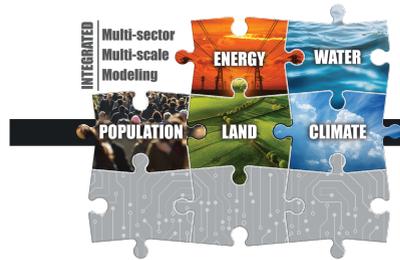
+

PLEXOS

=

nature

How Might MSD-LIVE Benefit You?



MOSART-WM

+

GCAM

+

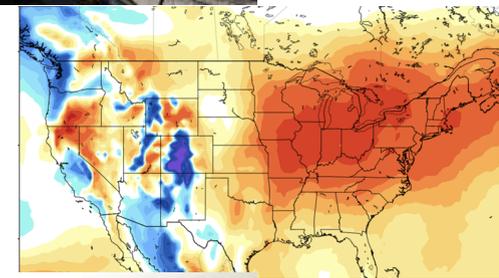
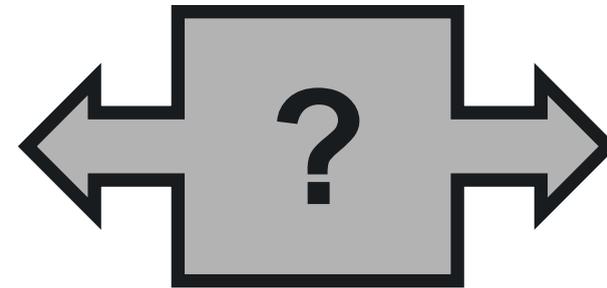
BEND

+

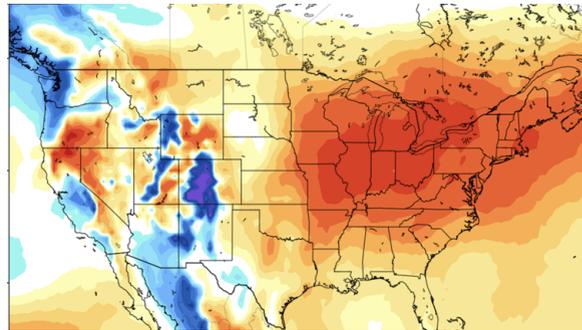
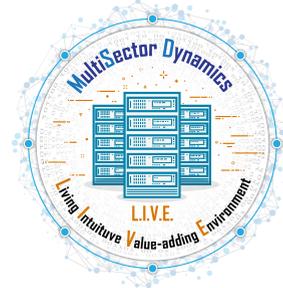
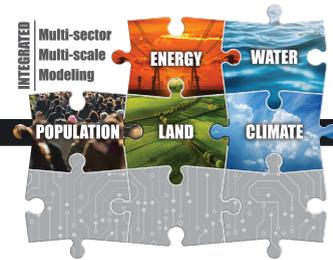
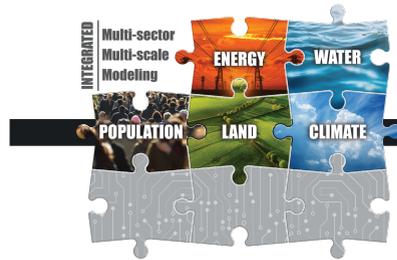
PLEXOS

=

nature



How Might MSD-LIVE Benefit You?



MOSART-WM

+

GCAM

+

BEND

+

PLEXOS

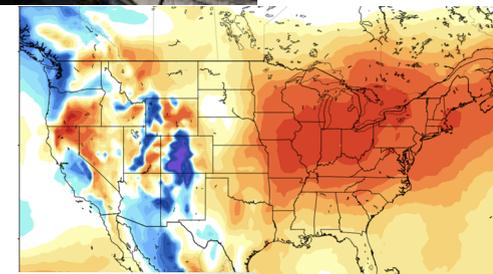
=

nature



Retrieve

- Input data + metadata
- Versioned model source code
- Initial and boundary conditions
- Data transformation scripts
- Model outputs

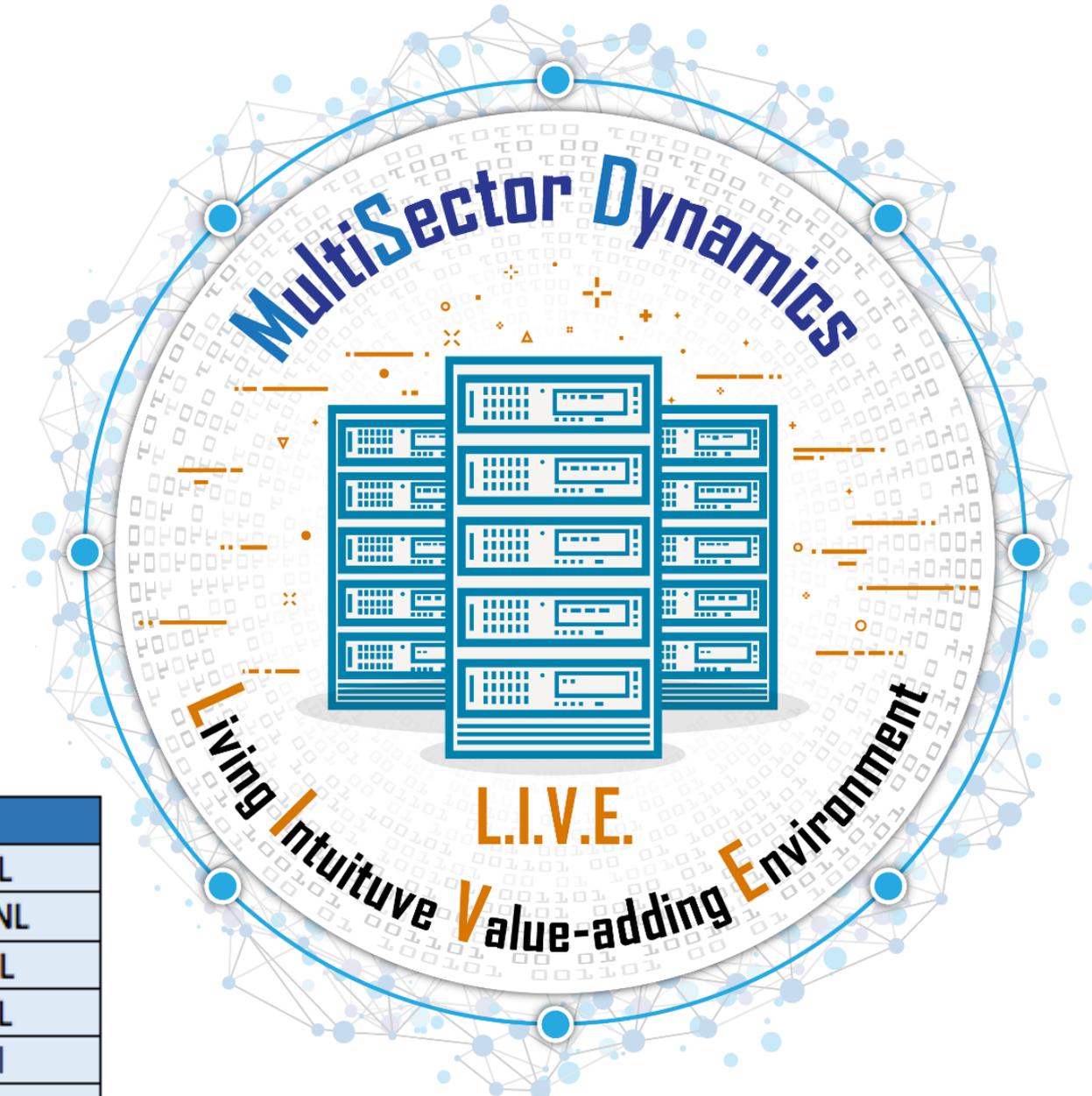


Project Designed with a Multi-Phase Implementation Plan

- **Phase 1** (Oct 2018 – May 2019):
 - Review of other archives, including ESS-DIVE, ARM, NCAR, and NASA platforms
 - Identification of user needs
 - Initial technical scoping and feasibility assessment
 - **White paper to BER in May 2019**
- **Phase 2** (May 2019 – Fall 2019):
 - Prototype and beta test core components
 - Anticipate a stakeholder meeting in fall 2019

*****Proposed Stakeholder Committee*****

Project, Person, Institution	Project, Person, Institution
IM3, Jennie Rice, PNNL	ESS-DIVE, Charu Varadharajan, LBNL
IM3, Maoyi Huang, PNNL	ARM and A2E, Chitra Sivaraman, PNNL
IHESD, Mohamad Hejazi, PNNL	OpenEI Repository, Jon Weers, NREL
IHESD, Robert Link, PNNL	IDEAS Project, David Moulton, PNNL
PCHES, Karen Fisher-Vanden, PSU	MSD Community, Pat Reed, Cornell
PCHES, Robert Nicholas, PSU	Coastal Project?
MIT, John Reilly, MIT	Arctic Project?



Project Designed with a Multi-Phase Implementation Plan

- **Phase 1** (Oct 2018 – May 2019):
 - Review of other archives, including ESS-DIVE, ARM, NCAR, and NASA platforms
 - Identification of user needs
 - Initial technical scoping and feasibility assessment
 - **White paper to BER in May 2019**
- **Phase 2** (May 2019 – Fall 2019):
 - Prototype and beta test core components
 - Anticipate a stakeholder meeting in fall 2019
- **Develop full proposal for BER** (spring 2020)
- **Phase 3** (FY20+ contingent on approval of full proposal):
 - Implementation of MSD-LIVE
 - Expand to MSD community outside of DOE

