

The logo for I-GUIDE, featuring the text "I-GUIDE" in a bold, white, sans-serif font against a dark blue background with a subtle map-like texture.The National Science Foundation (NSF) logo, featuring a blue globe with the letters "NSF" in white, surrounded by a golden sunburst pattern.

Institute for
Geospatial Understanding
through an Integrative
Discovery Environment

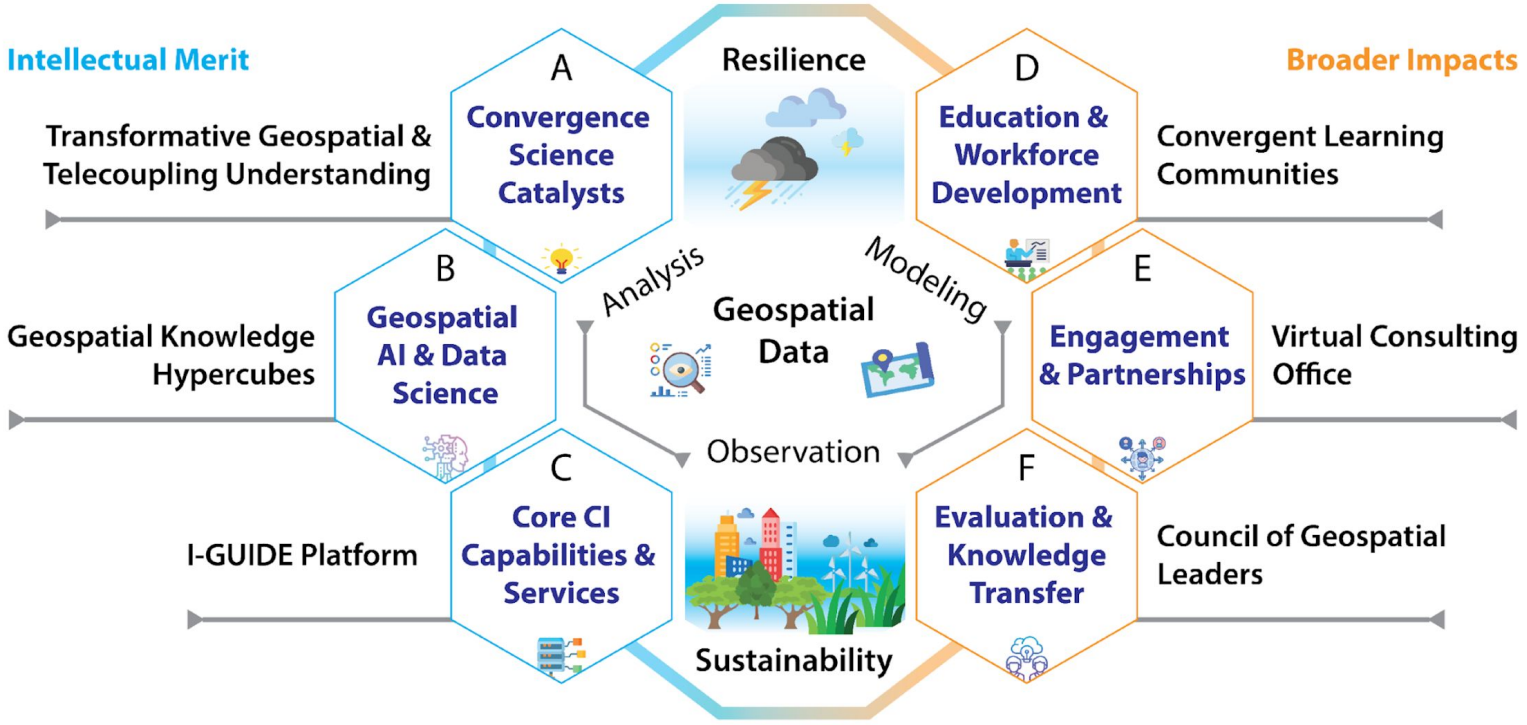
Carol X. Song, Ph.D., Chief Scientist, Rosen Center
for Advanced Computing, Purdue University, HDR
Institute for Geospatial Understanding through an
Integrative Discovery Environment

<http://www.rcac.purdue.edu/about/staff/cxsong>



Intellectual Merit

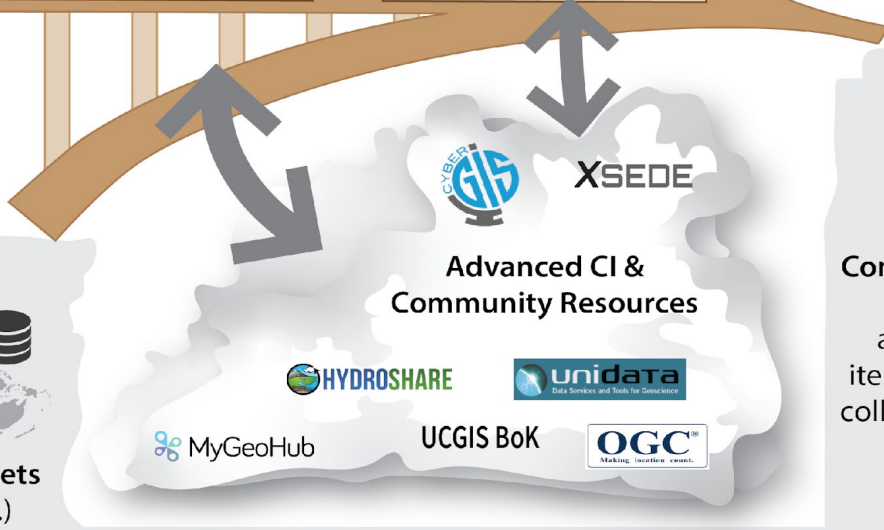
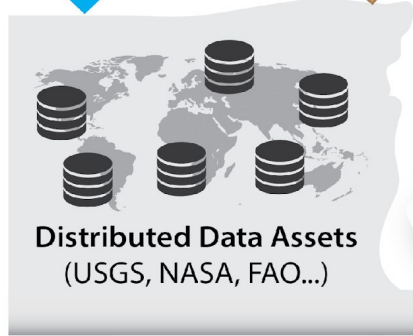
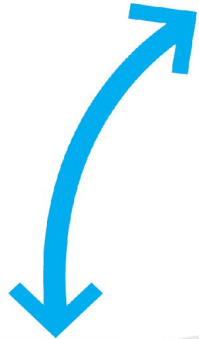
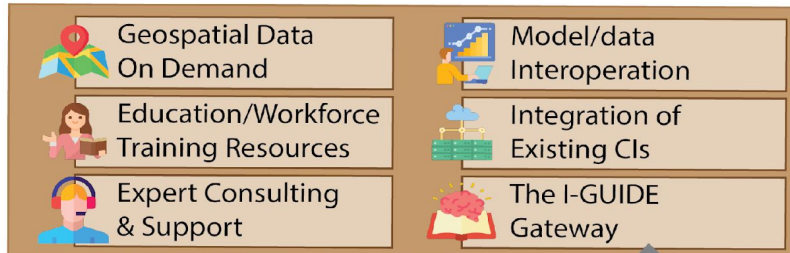
Broader Impacts



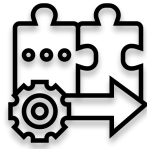


The I-GUIDE Platform (The “missing middle”)

Discovery & Solutions from
Convergence Research



Data/cyberinfrastructure best practices



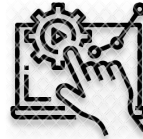
Integrate & Interoperate

Bring together disparate tools and datasets

Make access to HPC and cloud resources transparent

Make data actionable

Make workflows reproducible



Make Data/CI Useful and Usable

Data wrangling #1 challenge

Driven by pilots

Design for general applications via building blocks

Adopt familiar interfaces (e.g. notebooks) to lower barriers



Know Our Audience

Convergence researchers

Education and training activities

Broader engagements

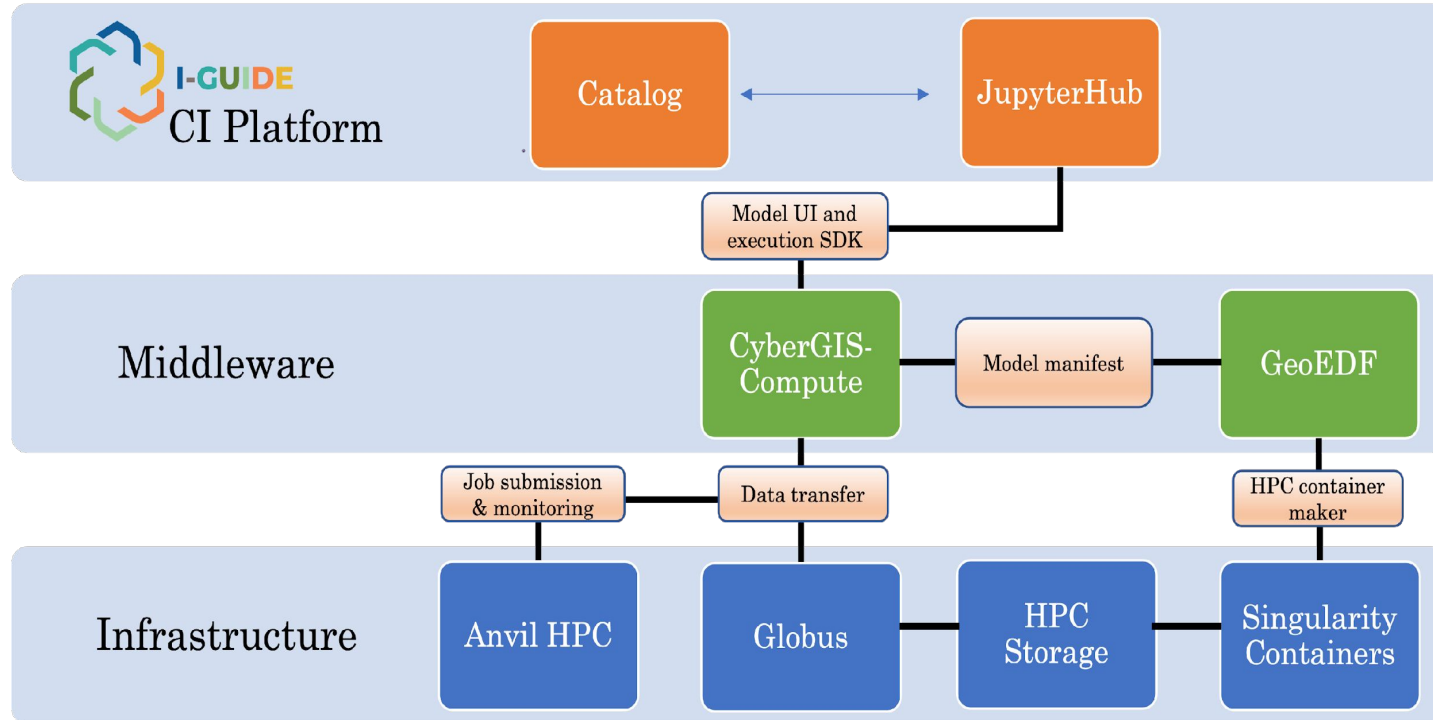
Knowledge transfer

Virtual consulting office

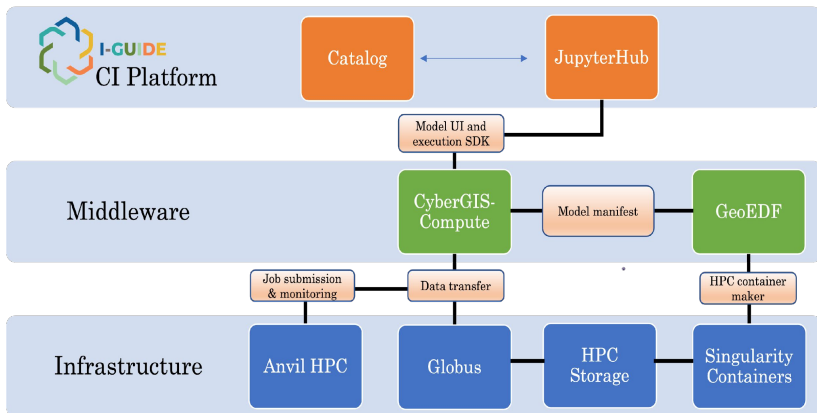
Data/cyberinfrastructure best practices



The I-GUIDE Platform integrates CyberGIS Compute, GeoEDF, Hydroshare, HPC & storage, Singularity, Globus, etc, into a seamless platform with Jupyter as the entry point.



Data/cyberinfrastructure best practices



```

iguide_wrfhydro.ipyb x iguide_agingsdams.ipyb x Python [conda env:iguide]
Code
benchmark_area = dam_area.loc[dam_area['Class'] == 0]
inund_area = dam_area.loc[dam_area['Class'] > 0]
benchmark_area_union = benchmark_area.dissolve(by='Dam_ID')
inund_area_union = inund_area.dissolve(by='Dam_ID')
fed_dams_focus = fed_dams.loc[fed_dams['ID'] == dam_id.reset_index()]

# Plot maps
var_list = ['no_diploma', 'poverty', 'less_english', 'mobile_home', 'no_vehicle', 'unemployment', 'age65']
for n in range(8):
    if n == 0:
        fed_dams_focus.plot(ax=ax[n], markersize=500, color='black')
        axin.set_title(label='Inundated Area', fontsize=24)
    else:
        dam_area.loc[dam_area[var_list[n-1]] != -666666666].plot(var_list[n-1], ax=ax[n], scheme='fisher')
        dam_area.loc[dam_area[var_list[n-1]] == -666666666].plot(ax=ax[n], color='Grey')
        axin.set_title(label=f'({var_list[n-1]}) (fed_dams_focus[\'c.\'] = var_list[n-1].values[0])', font=
        benchmark_area_union.boundary.plot(ax=ax[n], color='black', lw=0.5, linestyle='dashed')
        inund_area_union.boundary.plot(ax=ax[n], color='black', lw=1)
        axin.get_axis().set_visible(False)
        axin.get_yaxis().set_visible(False)

plt.show()
    
```

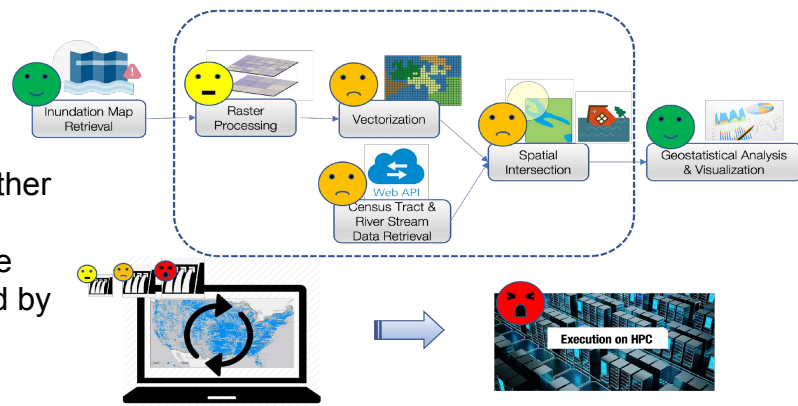
```

Data Visualization
[16]: for dam_id in fed_dams['ID'].unique():
      plot_vulnerability_attribute_around_dam(dam_id)

/opt/conda/envs/iguide/lib/python3.8/site-packages/geopandas/plotting.py:693: UserWarning: The GeodataFrame
you are attempting to plot is empty. Nothing has been displayed.
warnings.warn
/opt/conda/envs/iguide/lib/python3.8/site-packages/geopandas/plotting.py:693: UserWarning: The GeodataFrame
you are attempting to plot is empty. Nothing has been displayed.
warnings.warn
/opt/conda/envs/iguide/lib/python3.8/site-packages/geopandas/plotting.py:693: UserWarning: The GeodataFrame
you are attempting to plot is empty. Nothing has been displayed.
warnings.warn
/opt/conda/envs/iguide/lib/python3.8/site-packages/geopandas/plotting.py:693: UserWarning: The GeodataFrame
you are attempting to plot is empty. Nothing has been displayed.
warnings.warn
    
```



Address the questions of whether and where socioeconomically disadvantaged populations are vulnerable to disasters caused by potential aging dam failures.





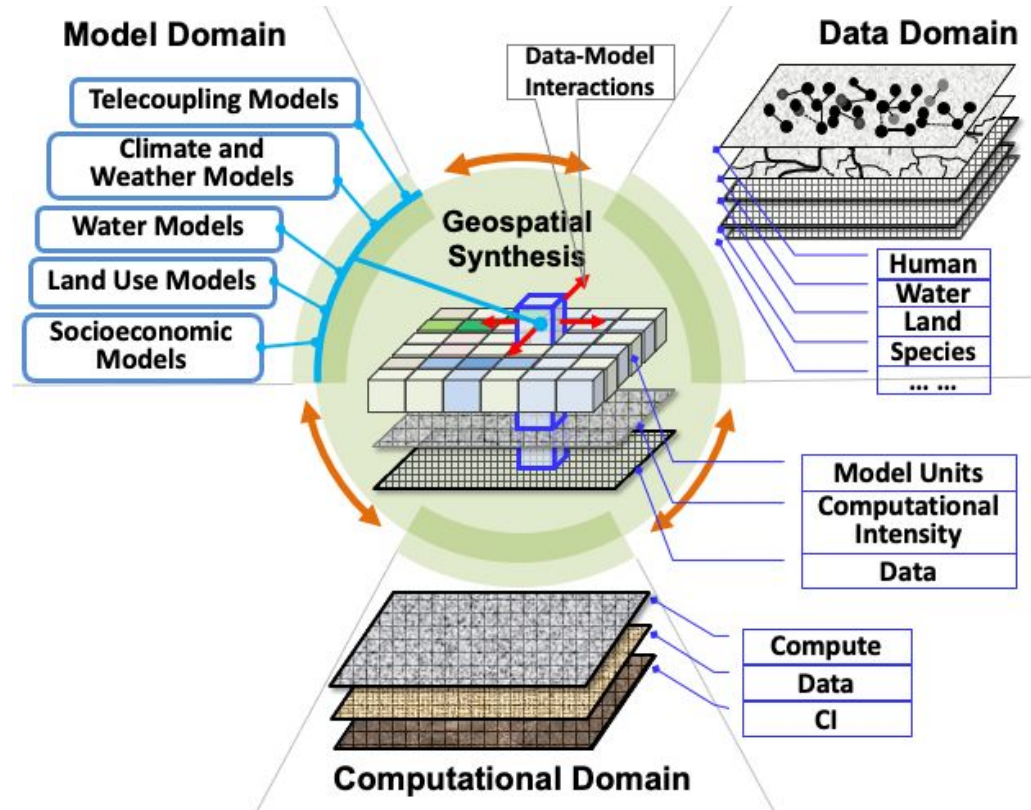
Impact: HDR and larger cyberinfrastructure

- Data wrangling tools brought together, especially for geospatial and temporal data
- A seamless platform that can be deployed by other projects
- A comprehensive resource catalog (actionable data) built on standards and interoperability
- Data hypercube to meet the challenges of hyper-dimensional data --- transforming data into knowledge
- The human factors - knowledge and best practices of building data analytics and computational cyberinfrastructure for convergence research

Current data/cyberinfrastructure needs



The overwhelming wealth of heterogeneous geospatial data presents significant challenges and opportunities to analytics and data-driven modeling approaches.



Current data/cyberinfrastructure needs



- Large scale data + computation + complexity (geospatial and temporal scales)
 - Lots of data wrangling required (heterogeneity of geospatial data)
 - Complex models, computation, analysis and data need to be linked
 - Large datasets needed for models (e.g., 150TB for one study) – how to make it available to broader communities
- Hyper-dimensional data
- Incorporating advances in geospatial AI and data science into CI and and make them usable beyond computer science labs
- Varying levels of CI expertise, gaps in cyber training