Building a World

The Data Challenges Associated with Low-Latency Environment Generation for Decision Support

Alexx Perloff[1], Michael Schneider[1], Jane Pratt[1], Dan Merl[2]

(1) Astronomy and Astrophysics Analytics Group, Physics Division, LLNL

(2) Machine Intelligence Group, Center for Applied Scientific Computing, LLNL

October 25, 2023

Decision Support For Decision Superiority

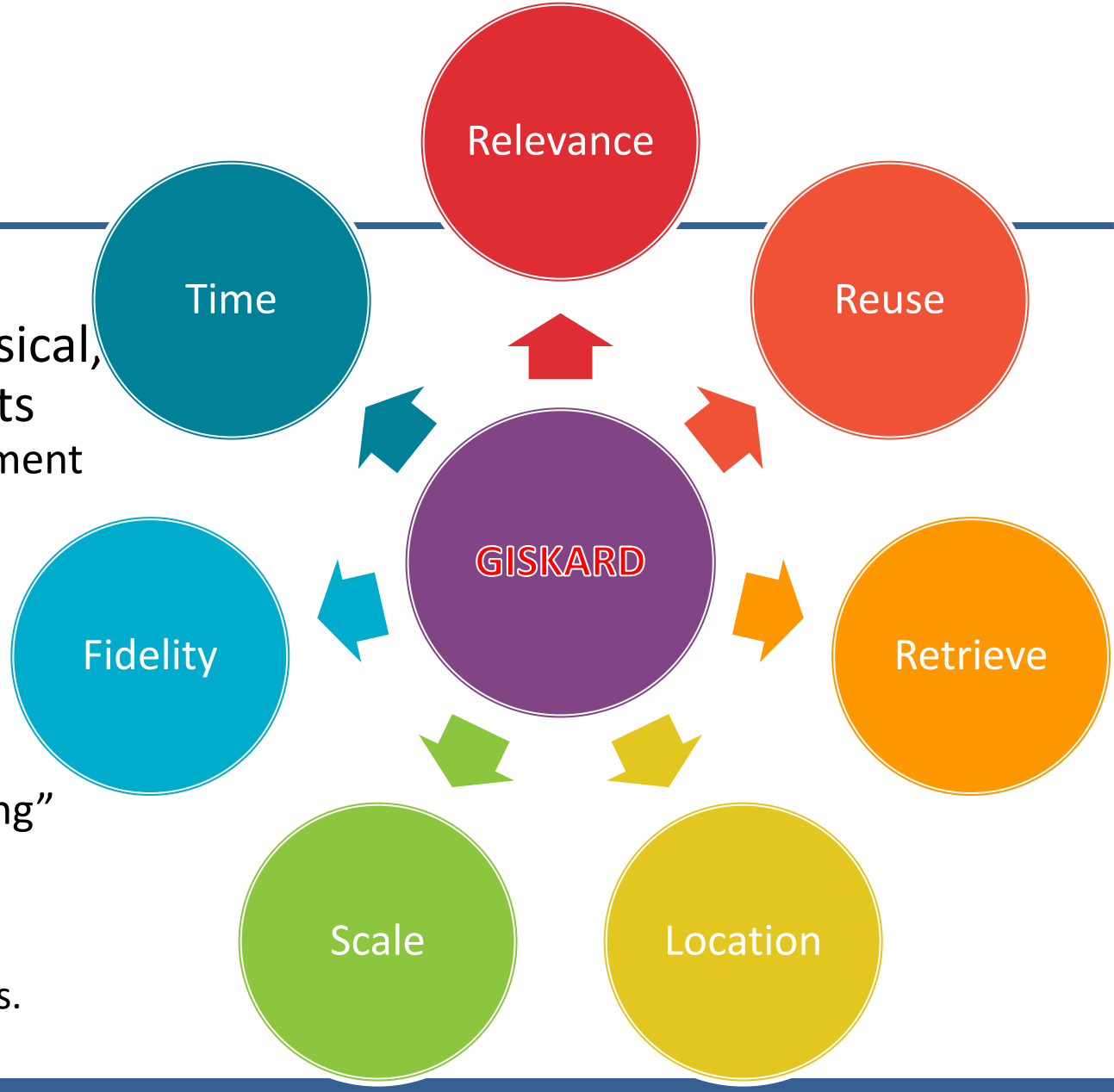
- Big changes in modeling combat environments can lead to big changes in how we achieve warfighting objectives
- Information dominance is revolutionizing warfighting and may also revolutionize wargaming
  - These tasks often entail collecting, organizing, storing, and sometimes moving vast amounts of information
- Decision superiority is enabled by optimizing conflict dynamics to achieve operational or strategic objectives
  - Human decision makers require computer-generated insights to cope with shrinking timelines and increasing complexity
- Combine novel computational planning approaches (machine learning) with conflict simulations on LLNL high-performance computing, informed by real-time data feeds

Graphic from: https://kriegsspiel.org/how-i-fell-in-love-with-kriegsspiel/

To achieve information dominance, we introduce data-driven dynamic maps over which we can optimize plans using computers

Project Goals

- Computer generated **decision support** requires the computer to understand physical, sociological, and environmental constraints
  - Build a high-resolution, dynamic map/environment for anywhere in the world
- Focus on the **data**
  - Data in … data out
- Focus on **environment synthesis**
  - Tie into multiple simulators for “decision making”
  - External plugins can…
    - be more extensible
    - be used with a wide array of simulators
    - focus on different features than current simulators.



Geospatial **I**nformation **S**ystem for **K**nowledge **A**nd **R**apid **D**ecisions (GISKARD)

Layers of Abstraction

Conflict Simulators

- JCATS/JLOD
- AFSIM
- EADSIM
- ABMARL
- PADL
- …

Mapping Programs

- Esri ArcGIS
- QGIS
- PostGIS
- Generic Mapping Tool
- …

Tools

- (Arch)GDAL
- GEOS
- Shapely
- …

Data Sources

- USGS
- NASA
- NOAA
- FAA
- GEBCO
- OpenTopography
- ESA
- ArcGIS Hub
- US Census Bureau
- …

Layers of Abstraction

Access Restrictions
Requires Login
API Discoverable
For Purchase
Never used
Somewhat familiar
Lots of experience

Conflict Simulators

- JCATS/JLOD
- AFSIM
- EADSIM
- ABMARL
- PADL
- …

Mapping Programs

- Esri ArcGIS
- QGIS
- PostGIS
- Generic Mapping Tool
- …

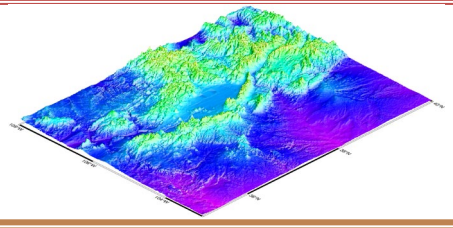
Tools

- (Arch)GDAL
- GEOS
- Shapely
- …

Data Sources

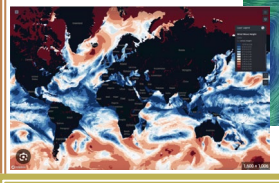
- USGS
- NASA
- NOAA
- FAA
- GEBCO
- OpenTopography
- ESA
- ArcGIS Hub
- US Census Bureau
- …

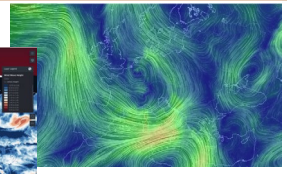
What makes an “environment”?



Terrain

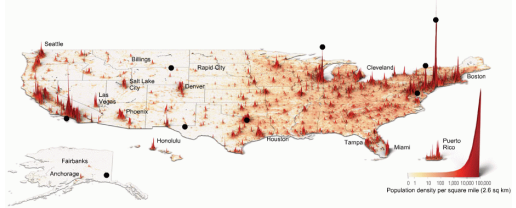
- Elevation (Land)
- Depth (Ocean)
- Building height





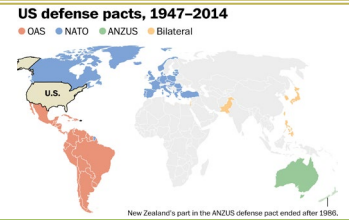
Atmosphere

- Wind
- Cloud Coverage
- Extreme Weather
- Precipitation



Regional Properties

- Demographics
- Monetary
- Political



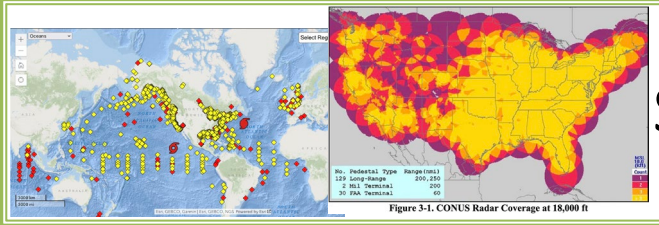
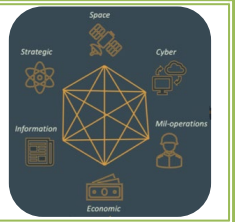


Figure 3-1. CONUS Radar Coverage at 18,000 ft

Sensors

- Radar
- Buoys
- Other sources of information



*All values are approximate

Big-data Operations

What does this all cost?

Elevation

- US rasters: ~28 TB
  - 1 meter (partial CONUS)
  - 5 meter (Alaska-only)
  - 1/9 arc-second (CONUS)
  - 1/3 arc-second (CONUS)
  - 1 arc-second (US + Canada)
  - 2 arc-second (Alaska-only)
  - 15 arc-second (global)
  - 30 arc-second (global)
  - 1 arc-minute (global)
- Global rasters (extrap.): ~471 TB
  - Assumes US ~6% land mass
- Format: GeoTIFF

Atmosphere

- **Global Data Assimilation System**
  - 0.25°~18 GB/day
  - 1.00°~1.8 GB/day
- **Global Forecast System Analysis**
  - 0.25°~414 GB/day
  - 0.50°~124.2 GB/day
  - 1.00°~41.4 GB/day
- **Two weeks of data:** ~8.2 TB
- **Format:** GRIB2
- **Updates:** 4 / day

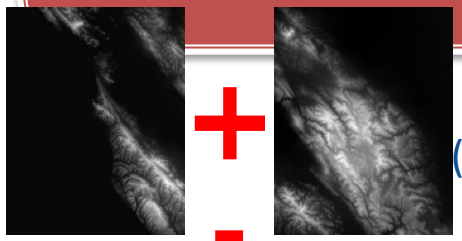
Demographics

- **Boundaries:** ~0 TB
- **Population:** ~0 TB
  - Country
  - State
  - County
  - Place/Remainder
- **Format:** Shapefile and CSV
- **Updates:** yearly

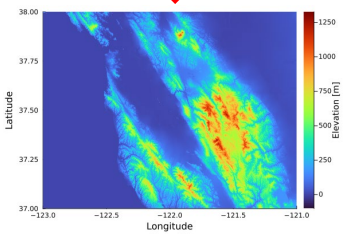
GIS at Scale

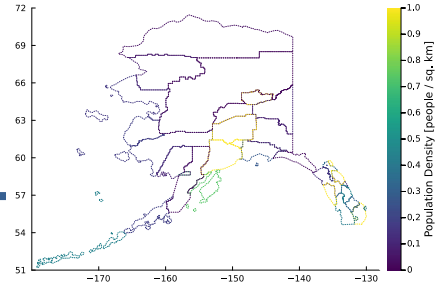
Elevation

- Arrange tiles
- Add placeholder tiles
- Down sample secondary dataset for missing tiles
- Select for region of interest (ROI)
- $\mathcal{O}(1\ \text{hour})$ at 1000 sq. deg.



Elevation
(California Bay Area)





Population Density
(Alaska)

Demographics

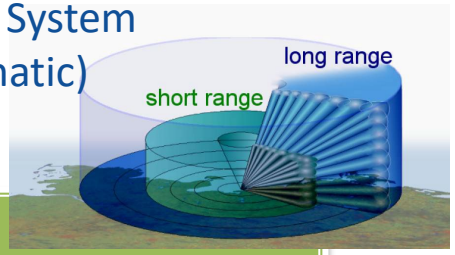
- Matching demographics to shapes
- Filter on ROI
- $\mathcal{O}(1\ \text{minute})$

North Warning System
(1987)

Sensors

- Join lists
- Filter duplicates
- Filter on ROI
- Generate shape based on capabilities
- $\mathcal{O}(1\ \text{minute})$

3D Radar System
(schematic)

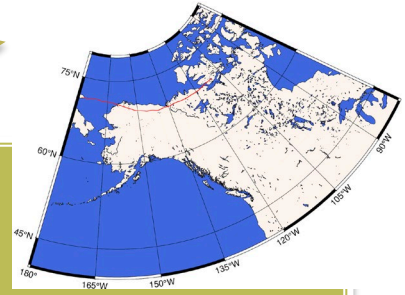


Scenario Builder

- Compute interaction among sensors/entities
- $\mathcal{O}(60\ \text{minutes})$ at 1000 sq. deg.

Atmosphere

- Propagate objects
- $\mathcal{O}(2\ \text{minutes})$

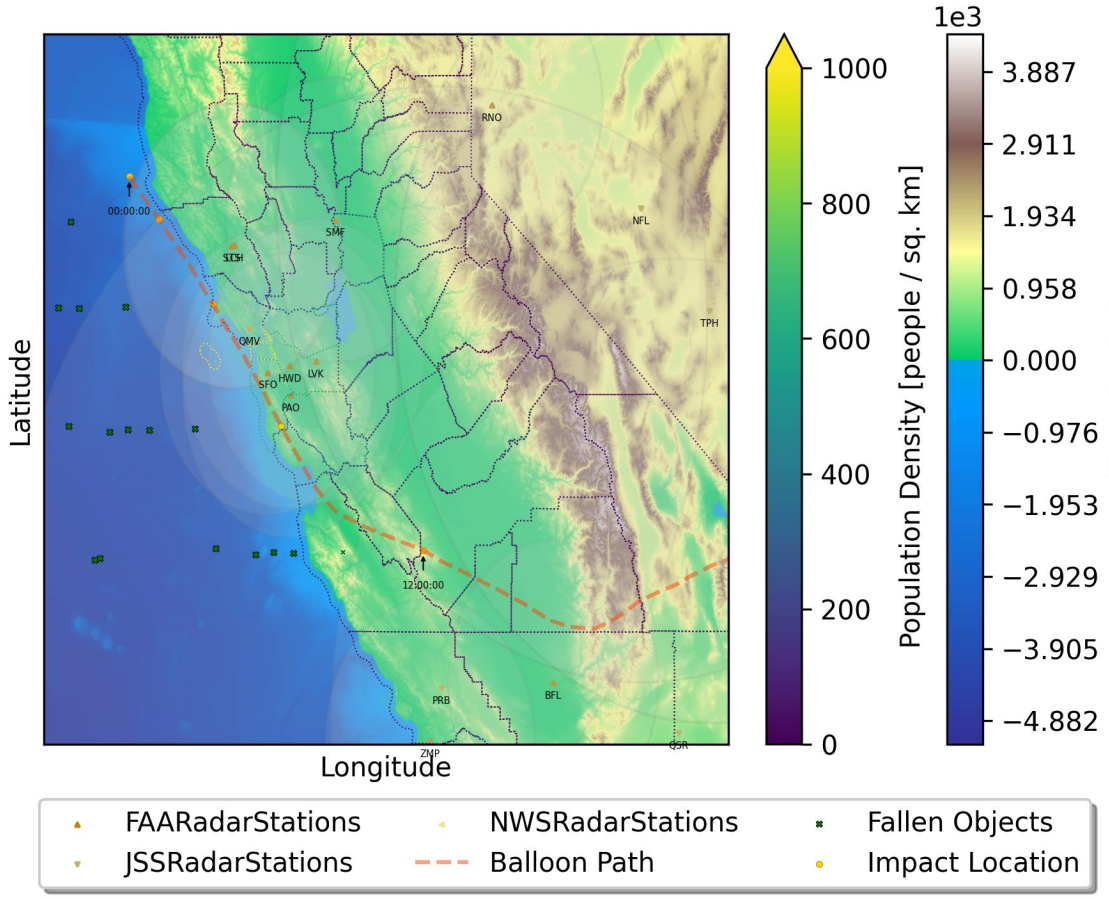


Balloon Propagation
(20230411T06 - 20230415T05)

From Environment Generator to Conflict Simulator

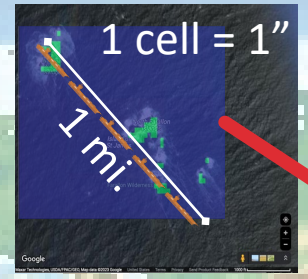
Teamwork makes the dream work

- **“Fast” environment creation**
  - Send map on simulator scenario initialization
    - Don’t want to be waiting hours/days to acquire and process relevant data
    - Don’t want the simulation to need to wait for GISKARD
- **Perform some basic analytics**
  - Example: post-engagement analysis
  - Send information to the simulator in addition to the environment
- **Completeness vs speed/automation**
  - 3D grids
  - Variable sized grids
  - Merging inputs at multiple latitudes
- **Scriptable**

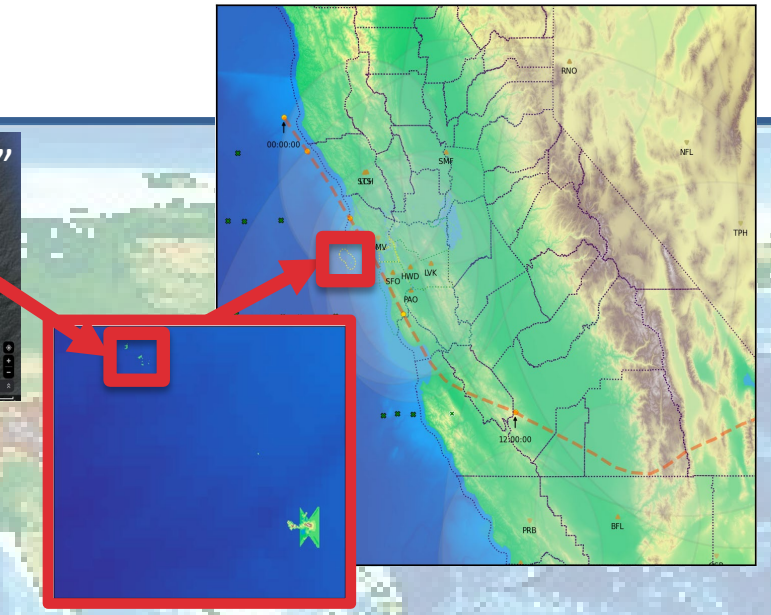


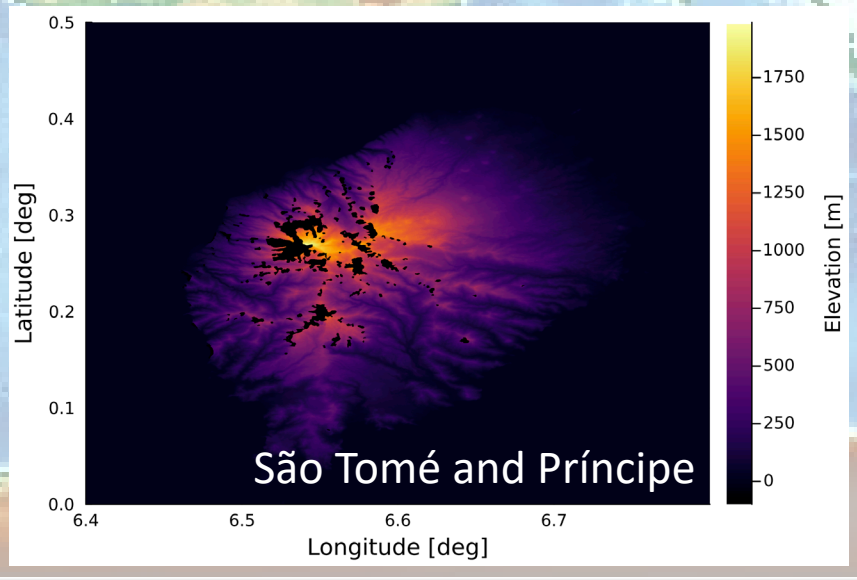
Computational Challenges

- Inputs have regular & irregular update schedules
- Lots of time downloading/caching data
- Large IO (especially for elevation rasters)
- Large memory footprint (especially for elevation rasters)
- Only some pieces parallelizable
- Always want safe operations
  - Data quality is a continual issue



28801” x 18001”



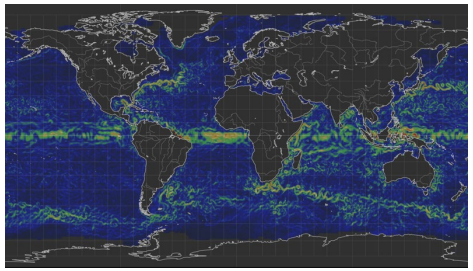


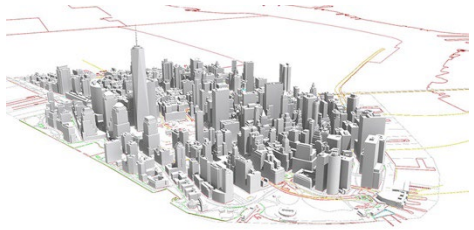
Tozer et al., 2019

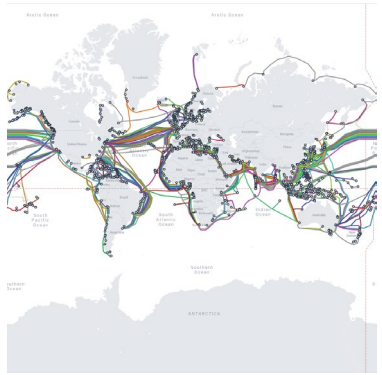
Looking to the Future

Source identified
Potential source / by-hand data collection
No source
Pending

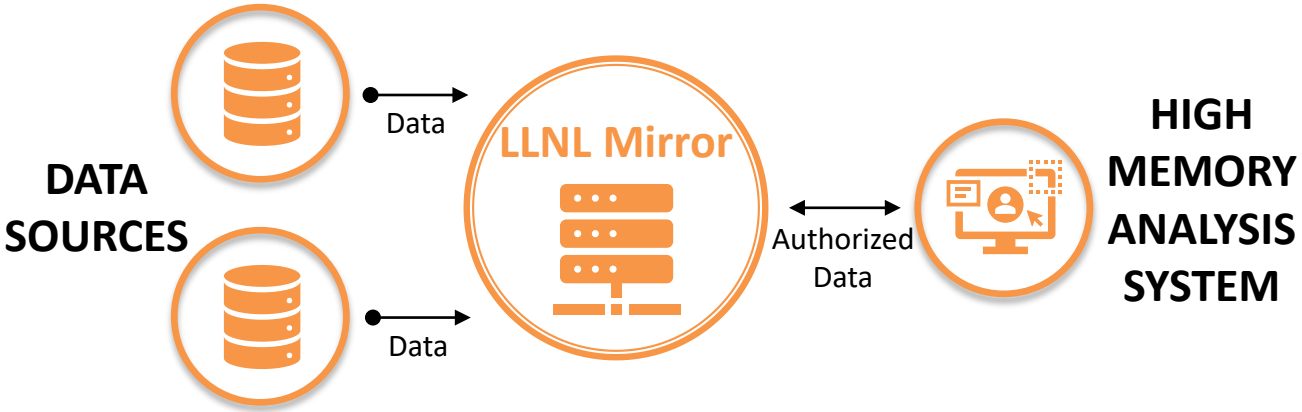
- **New sources of information**
  - Ocean currents
  - Cloud coverage
  - 3D building information
  - Terrain analysis
  - Information space (pseudo-geospatial)
  - Troop/resource positions
  - Classified data feeds







- **Reduced time to result**
  - Track changes to remote inputs
  - Pull in new data
  - Cache (fast disk access) for later analysis



- **Continually updated environment**
  - Bi-directional updates
    - Simulator request: on change in ROI
    - GISKARD push: on dynamic layer change
- **Preprocessing inputs**

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC