# geoKepler Workflow Module for Computationally Scalable and Reproducible Geo-Processing and Modeling

## Cyberinfrastructure for Wildfire Research and Response

Daniel Crawl[1], Jessica Block[2], John Graham[2], Charles Cowart[1], Amarnath Gupta[1], Mai Nguyen[1], Raymond de Callafon[3], Larry Smarr[2], Ilkay Altintas[1]

1-San Diego Supercomputer Center, UC San Diego, 2-California Institute for Telecommunications and Information Technology, UC San Diego

3-Mechanical and Aerospace Engineering, UC San Diego

**WIFIRE**
*wifire.ucsd.edu*

**NH43B-1887**

---

## What is WIFIRE

WIFIRE is an NSF-funded project to build an end-to-end cyberinfrastructure for real-time and data-driven simulation, prediction, and visualization of wildfire behavior. WIFIRE may be used by wildfire management authorities in the future to predict wildfire rate of spread and direction, and assess the effectiveness of high-density sensor networks in improving fire and weather predictions. WIFIRE has created a data model for wildfire resources including sensed and archived data, sensors, satellites, cameras, modeling tools, workflows, and social information including Twitter feeds for wildfire research and response.

### Programmable Workflows

**Kepler**
kepler-project.org

The Kepler scientific workflow system is instrumented to interact with the data catalog to access real-time streaming and archived wildfire data and stream it into dynamic data-driven wildfire models at scale.



---
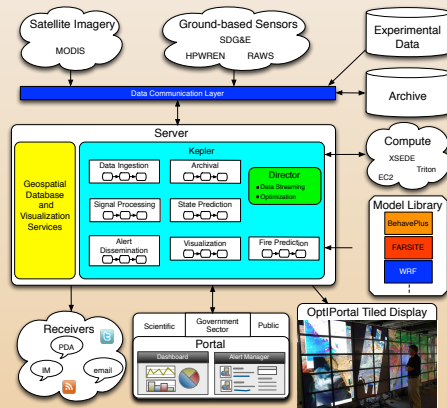
## What is geoKepler

| Data Access & Transfer | Geospatial Processing | Scalable Fire Modeling | Visualization |
|---|---|---|---|
| • WFS, REST<br>• Shapefile, GeoJSON, KML<br>• GeoTiff, Landscape | • Geotools, GDAL<br>• Reproject, Union, Filter, Rasterize | • FARSITE, WindNinja<br>• Spark, Hadoop<br>• R, Matlab | • Fire Perimeters<br>• Wind Vectors<br>• Photo Processing |

geoKepler includes a set of reusable GIS components, or actors, for the Kepler Scientific Workflow System (https://kepler-project.org/). Actors exist for reading and writing GIS data in formats such as Shapefile, GeoJSON, KML, and using OGC web services such as WFS. The actors also allow for calling geoprocessing tools in other packages such as GDAL and GRASS. Kepler integrates functions from multiple platforms and file formats into one framework, thus enabling optimal interoperability, model coupling, and scalability. Products of the GIS actors can be fed directly to models such as FARSITE and WindNinja. Kepler's ability to schedule and scale processes using Hadoop and Spark also makes geoprocessing ultimately extensible and computationally scalable.

The reusable workflows in geoKepler can be made to run automatically (as single executions or as many in parallel) when alerted by real-time environmental conditions. Here, we show breakthroughs in the speed of creating complex data for hazard assessments with this platform. We also demonstrate geoKepler workflows that use Data Assimilation to ingest real-time weather data into wildfire simulations, and for data mining techniques to gain insight into environmental conditions affecting fire behavior. Kepler components for data mining based on existing machine learning tools and libraries such as R and MLlib are being leveraged for this purpose, as well as Kepler's Distributed Data Parallel (DDP) capability to provide a framework for scalable processing.
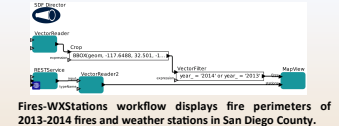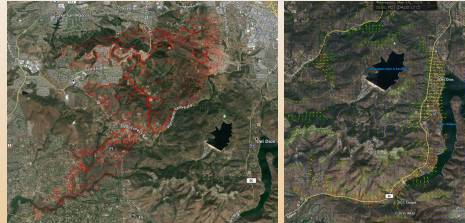
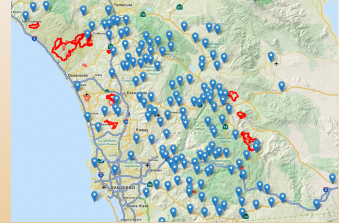### Example geoKepler Workflows



Santa Ana Conditions Workflow creates polygons of wind velocities and directions matching certain criteria.
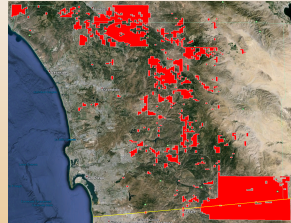


Fire Growth Workflow runs WindNinja and FARSITE.



Fires-WXStations workflow displays fire perimeters of 2013-2014 fires and weather stations in San Diego County.

---