

# Workflow Based Tools for Integrated Spatiotemporal Research

**Wendy Guan, Lingbo Liu & Xiaokang Fu**  
*Center for Geographic Analysis, Harvard University*  
**Shuming Bao**  
*Future Data Lab*

# The Era of the 4th Industrial Revolution

*Artificial Intelligence*

*Digital Twins*

*Cloud Computing*

*Big Data*

*Internet of Things*

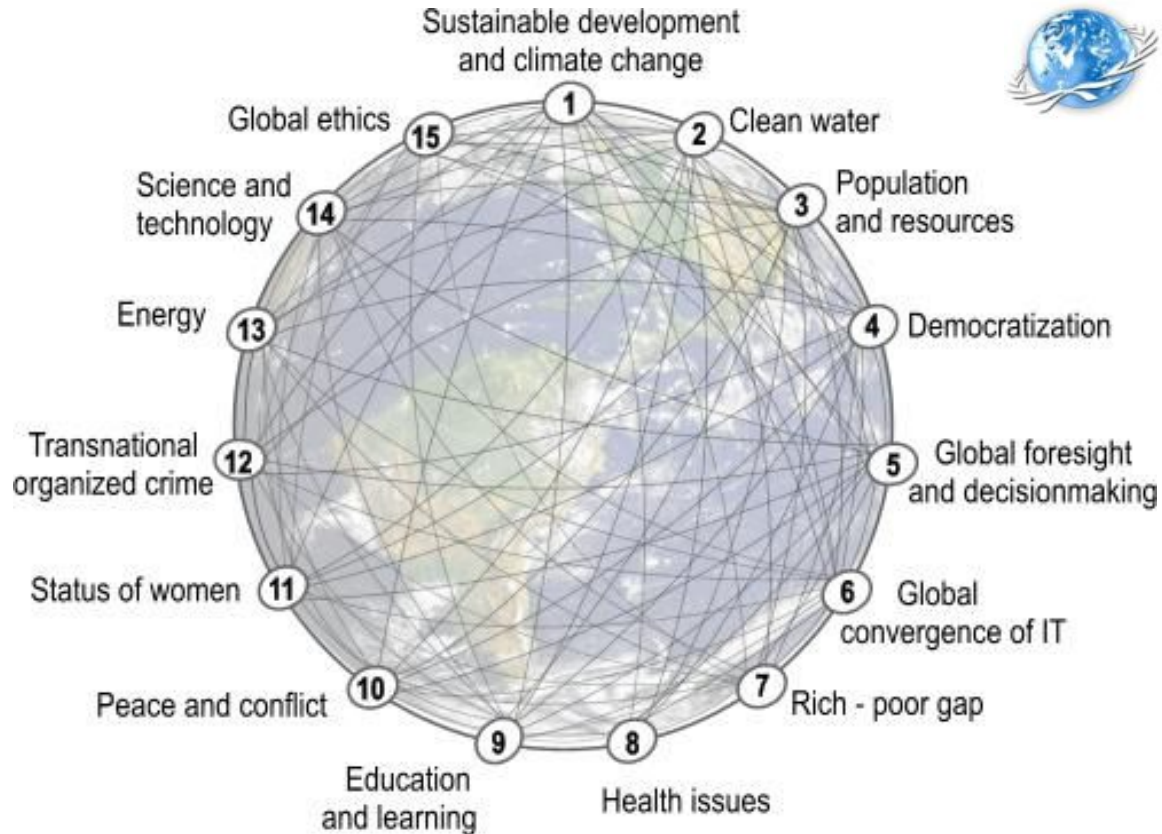
*Robotics*

*Bioinformatics*

*Block Chain*

*Geolocation becomes ubiquitous, place and time are embedded*

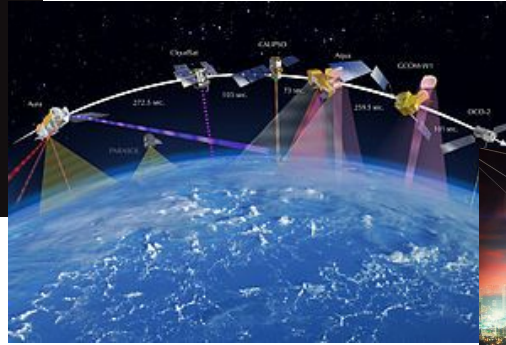
# The 15 Global Challenges



**The Millennium Project**  
GLOBAL FUTURES STUDIES & RESEARCH

***Overcoming these challenges requires the insight of complex systems on earth.***

# Geospatial & Temporal Data Are Everywhere



Global Navigation Satellite Systems (GNSS)

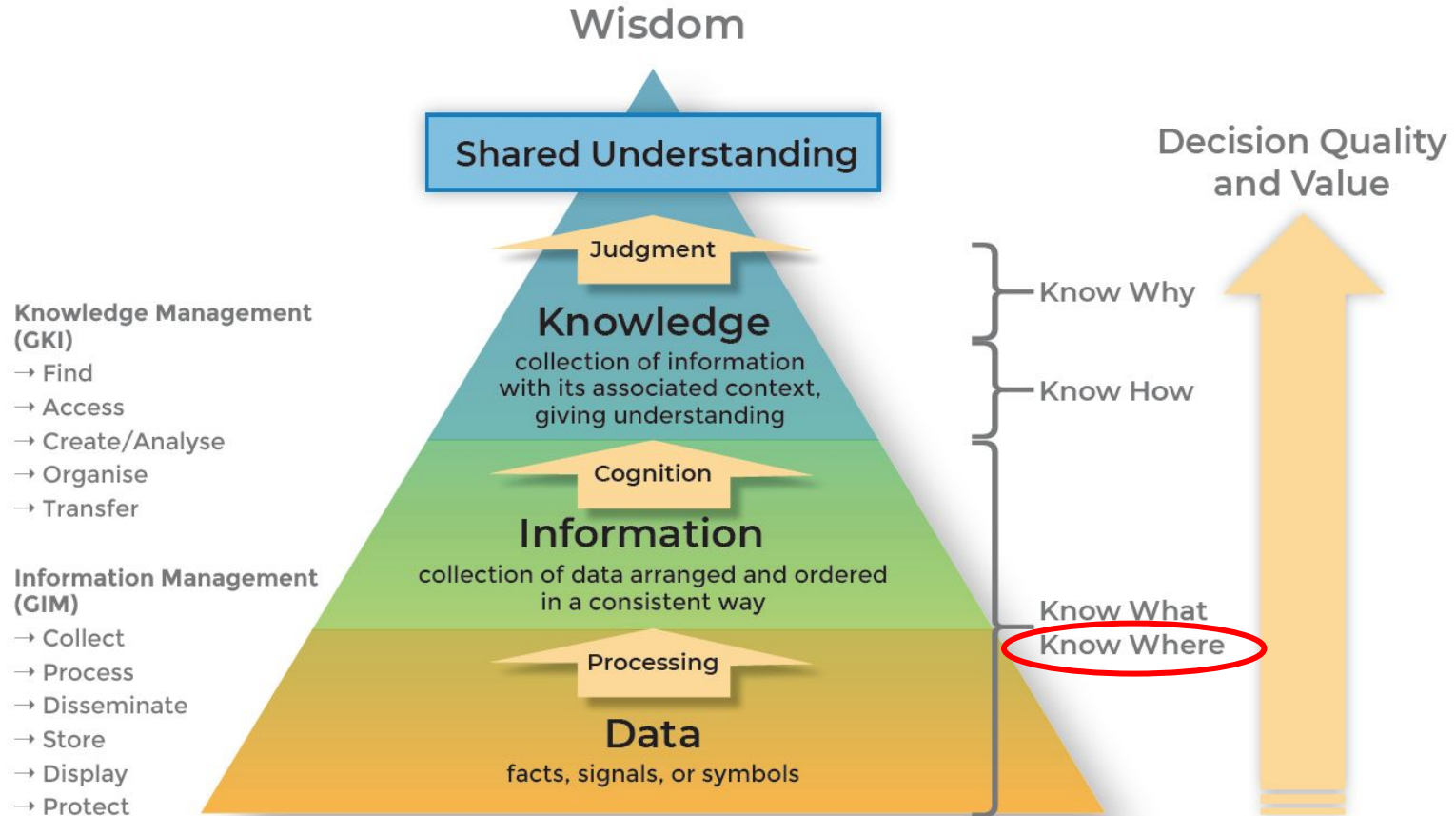
Earth Observation Satellites

Smart Phones, Clothes, Cars, Homes, Cities,

...



# How to Solve Problems with It?





# Spatial Data Science Research Needs

- Easy data access, especially for frequently updated data
- Heterogeneous data integration
- Efficient data analysis and visualization
- Access to high performance computing, especially for spatiotemporal big data
- Reproducible and replicable results
- Collaboration among researchers from different fields with different levels of skills



Image Source:

<https://www.sudeep.co/data-science/2018/02/09/Understanding-the-Data-Science-Lifecycle.html>

# Spatial Data Services

- Data & metadata download
- Query and pre-view
- Shared data editing
- Online Analysis
- Publishing
- Preservation



# Challenges in Spatial Data Services

- High development cost and slow implementation
- High maintenance requiring professional skills
- Lack of flexibility for customizable research inquiries and changing demands
- Barriers when sharing with researchers from different fields at different knowledge and skill levels



Image Source:  
[https://aub.edu.lb/libguides.com/data\\_services/home](https://aub.edu.lb/libguides.com/data_services/home)



# Challenges in Spatiotemporal Research

- Data Barriers
  - Discovery (sources unknown, scattered)
  - Accessibility (restricted, sensitive, licensed)
  - Big data (volume, variety, velocity, veracity)
- Tool Barriers
  - License
  - Operating environment
  - Computing power
  - Maintenance (security, version updates)
- Implementation Barriers
  - Methodology (reproducible, replicable, generalizable)
  - Technology (collaborators' varying backgrounds & skills)
  - Applications (efficiency, effectiveness, scalability)



**About 74% re-executions failed**

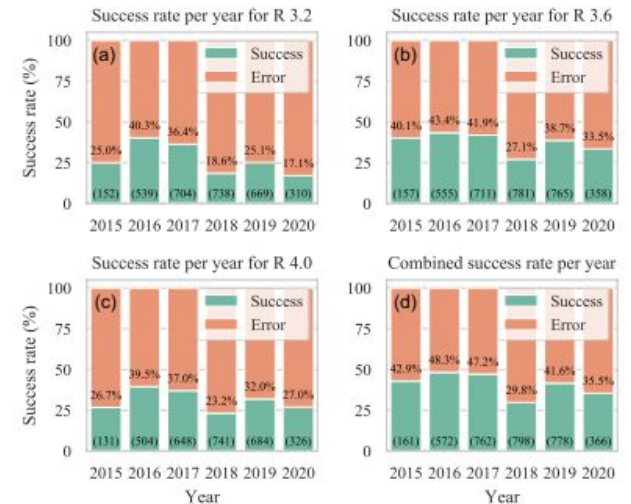


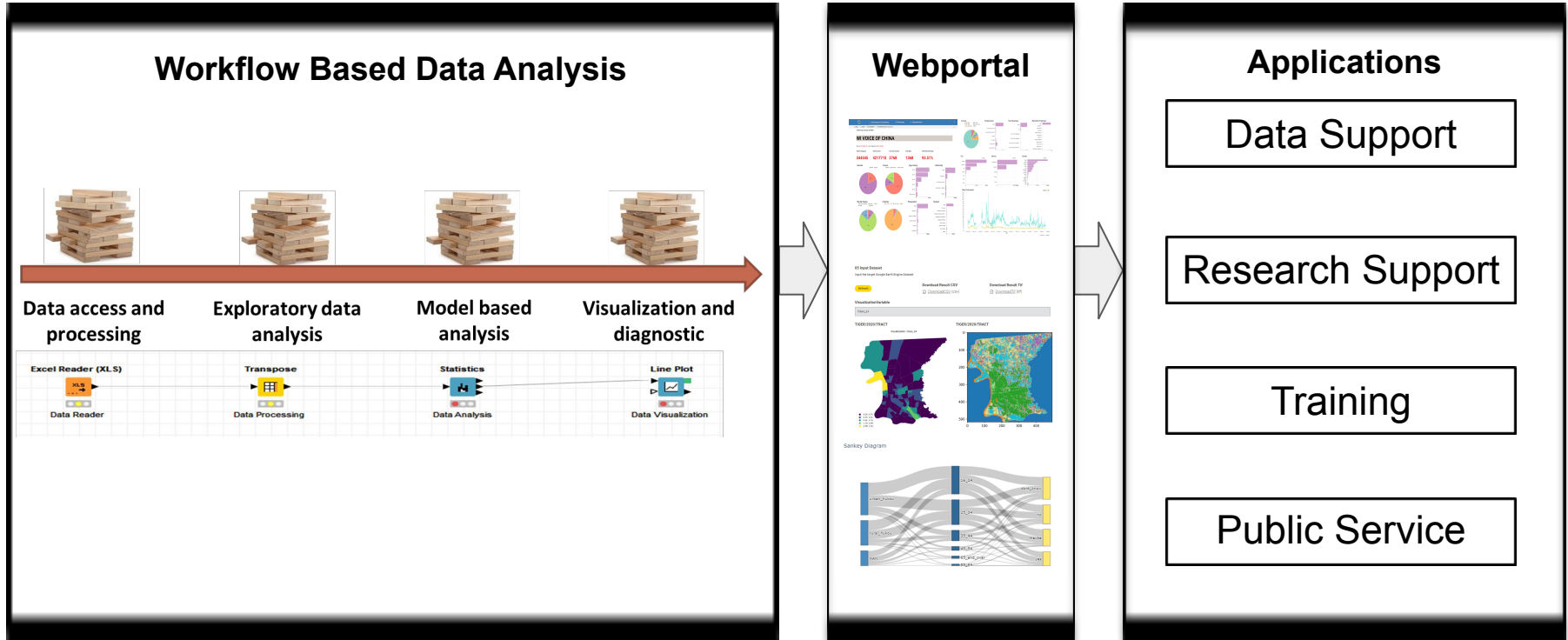
Fig. 11 Re-execution success rates per year per R software version.

Image source:

<https://www.nature.com/articles/s41597-022-01143-6.pdf?origin=ppub>

# Our Approach - the Spatial Data Lab (SDL)

An Integrated Solution for Spatiotemporal Research, Applications, Training, and Services



# About the Spatial Data Lab Project

Project in



NSF IUCRC  
Spatiotemporal Innovation  
Center(STC)



## GOAL

Build a platform for effective and collaborative spatiotemporal research

Sponsored by



Future Data Lab



## OBJECTIVES

Cloud-based spatial data integration



Executable quantitative analysis tools



Workflow-based case studies



Education and training

Conducted at



Center for  
Geographic Analysis

Harvard University

## MAJOR PRODUCTS

Spatial data services



Spatial data analysis platform and tools



Workflow based case studies



Training programs & research papers

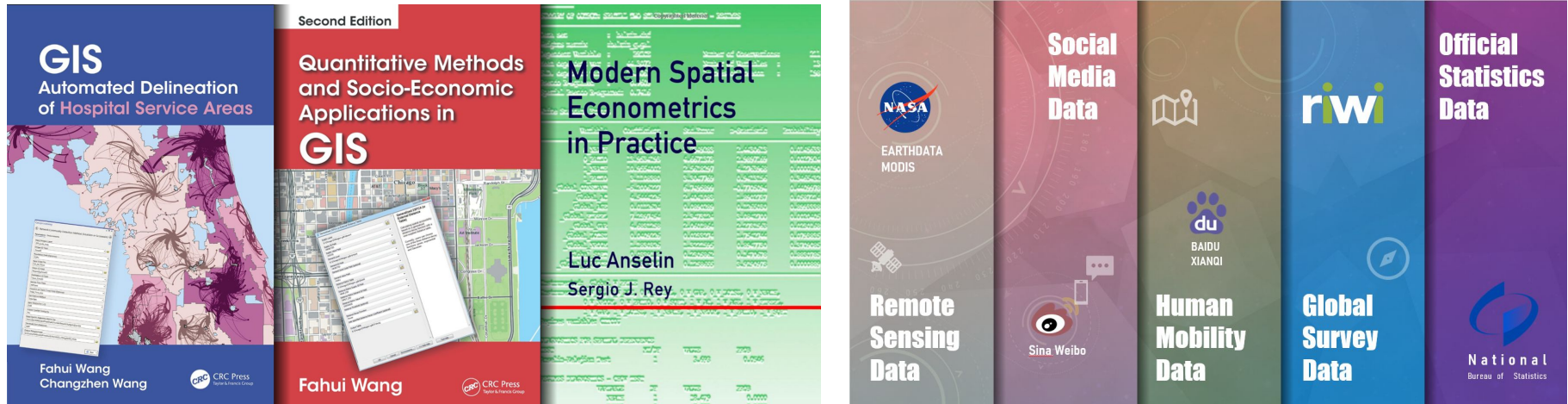
Hosted at



GEORGE  
MASON  
UNIVERSITY

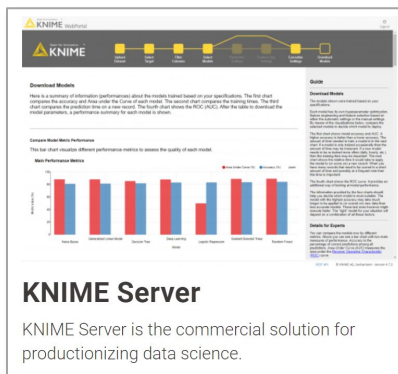
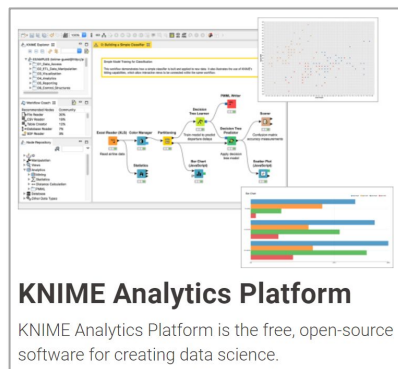
# Current and Previous Work

- **Data collection and sharing:** COVID-19 Data, Mobility Data, available via Dataverse API
- **Research applications and support:** Workflow-based Case studies ( 60+ case studies based on peer-reviewed publications; team members published 30+ papers)
- **Workbooks for textbook and teaching lab:** 30+ case studies (theory, models, and tools)
- **Training workshops:** 30+ training workshops on research data collections, COVID-19 studies, and spatiotemporal studies on various topics



# KNIME as the Foundation for the SDL

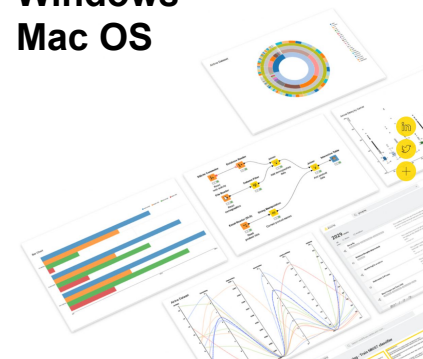
- Open source (<http://knime.com>)
- End to End Data Science, enabling automation from data input , manipulation, visualization, mining, reporting and web-based services
- Codeless , Visual and workflow-based Programming with 2000+ built-in nodes
- Extension for coding snippet of Python/ R /Java
- Integrated environment for desktop version and server version
- Workflow-driven Web Portal service for easy-to-use, interoperable, and executable data modelling and reporting



## End to End Data Science

At KNIME, we build software to create and productionize data science using one easy and intuitive environment, enabling every stakeholder in the data science process to focus on what they do best.

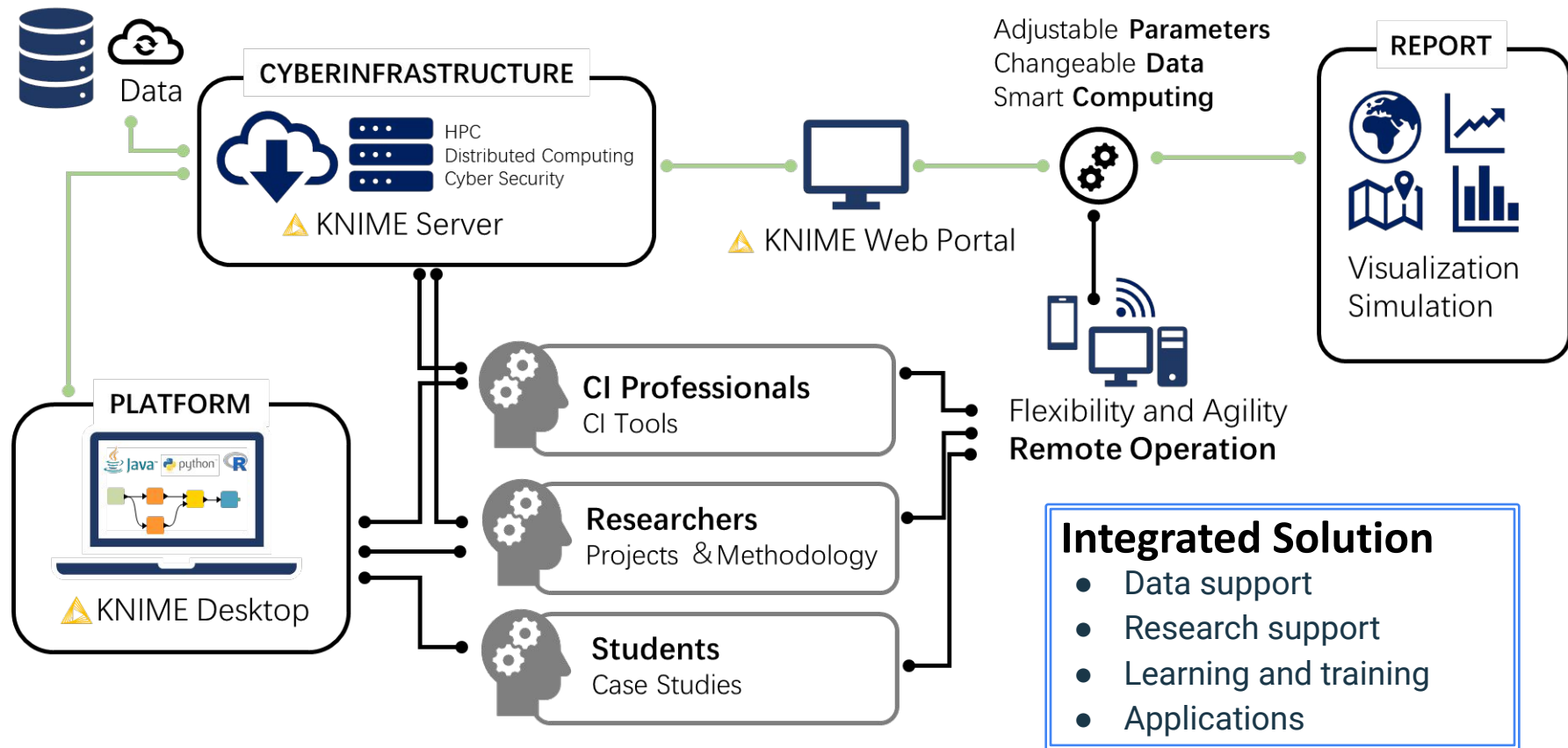
**Linux  
Windows  
Mac OS**



**Image Source:**  
<https://www.knime.com/>



# The KNIME-based SDL Architecture Overview



# The Integrated Solution Supporting Spatiotemporal Research

## Data Support

- Out-of-Box Application without Coding
- Adjustability on Key Parameters in Modelling
- Reproducible with New Data

## Research Support

- Integration of Local and Server for Sharing
- Data Privacy Protection
- Expandable Workflows

## Learning & Training Support

- Ubiquitous Access
- Flexible Skill Requirements
- Tailorable Modules

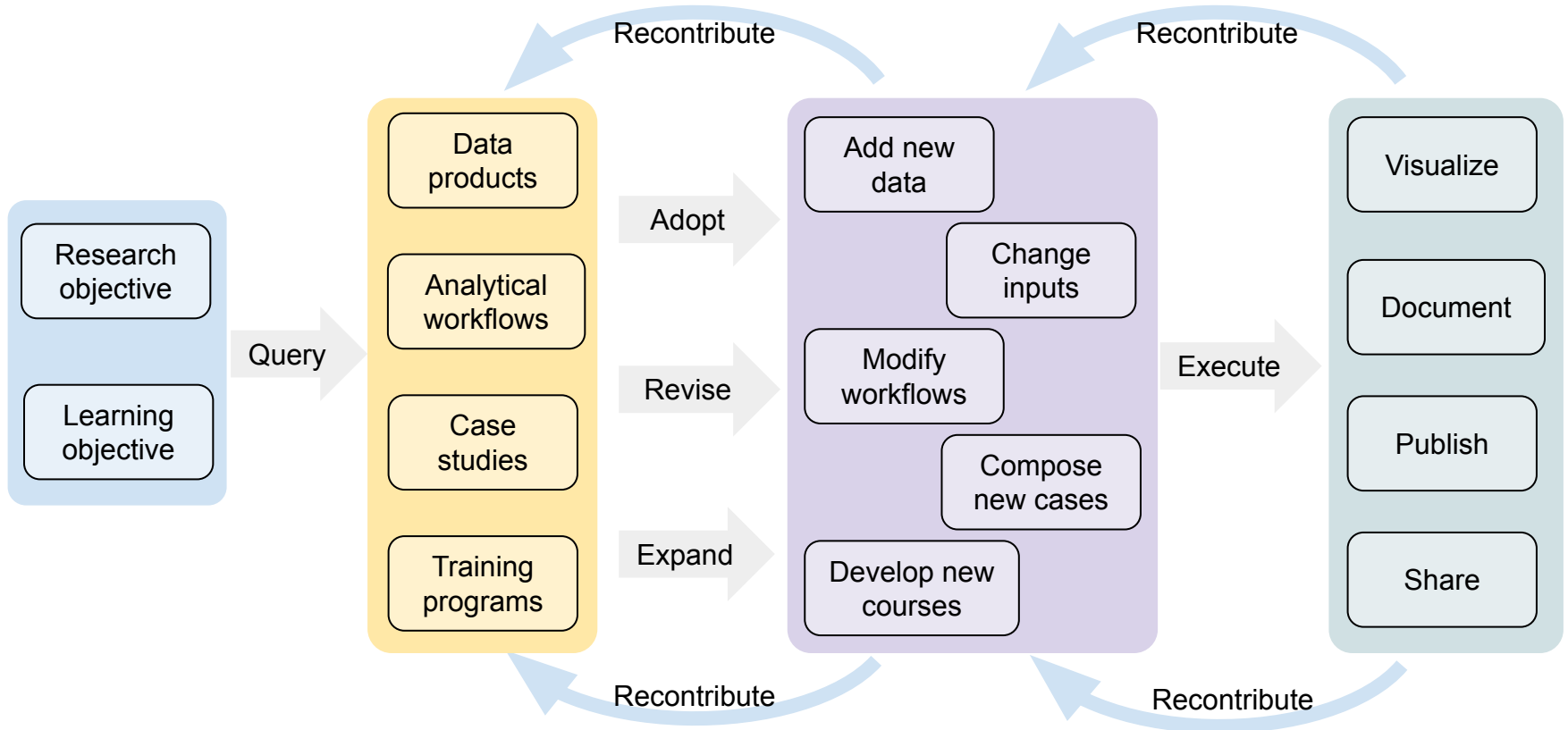
## Application Support

- Data and Report Automation
- Easy Maintenance
- Interactive Usage on Web Page

# Advantages in Spatial Data Service

- Efficient development and implementation of customized research services
- Easy maintenance of the system without special skills
- Great flexibility on expansion for changing demands
- Low entry cost for learning and applications
- Promoting reproduction and replication of quantitative research
- Facilitating multidisciplinary and cross-organizational collaboration

# The Re-engineered Research Lifecycle in SDL



# Geospatial Analytics Extension for KNIME

- Release in December 2022
- Easy installation via drag and drop from KNIME Community Hub
  - Batteries included - no additional setup and installation steps required
- Support for most common vector data, e.g. points, lines, polygons, collections
- Joined development by the Center for Geographic Analysis at Harvard and KNIME
- We will continue to work together to add more functionality and improve the usability
- Feedback is always welcome via KNIME Forum or Github\*

\*<https://github.com/spatial-data-lab/knime-geospatial-extension>



# End to End Geospatial Analytics



Spatial IO



Spatial Calculation



Spatial Manipulation



Spatial Transformation



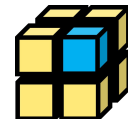
Spatial  
Conversion



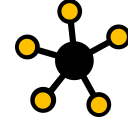
Spatial Visualization



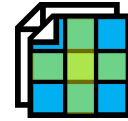
Exploratory Spatial  
Data Analysis



Spatial Modelling



Location Analysis



Open Datasets

# The Application Cases of Geospatial Analytics

3:10



Case 1

[Identification of Nearby Facilities](#)

3:38



Case 2

[Site Selection by Accessibility for Vaccination](#)

3:36



Case 3

[The Impact Assessment of Hurricane Ian](#)

3:30



Case 4

[Demographic Analysis with the US and China 2020 County Census Data](#)

4:08



Case 5

[Building an OSM POI Heat Map Dashboard](#)

3:38



Case 6

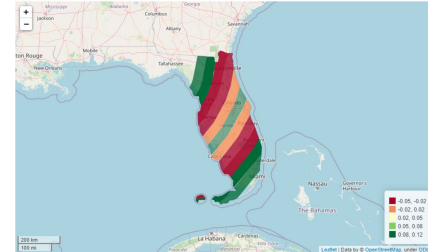
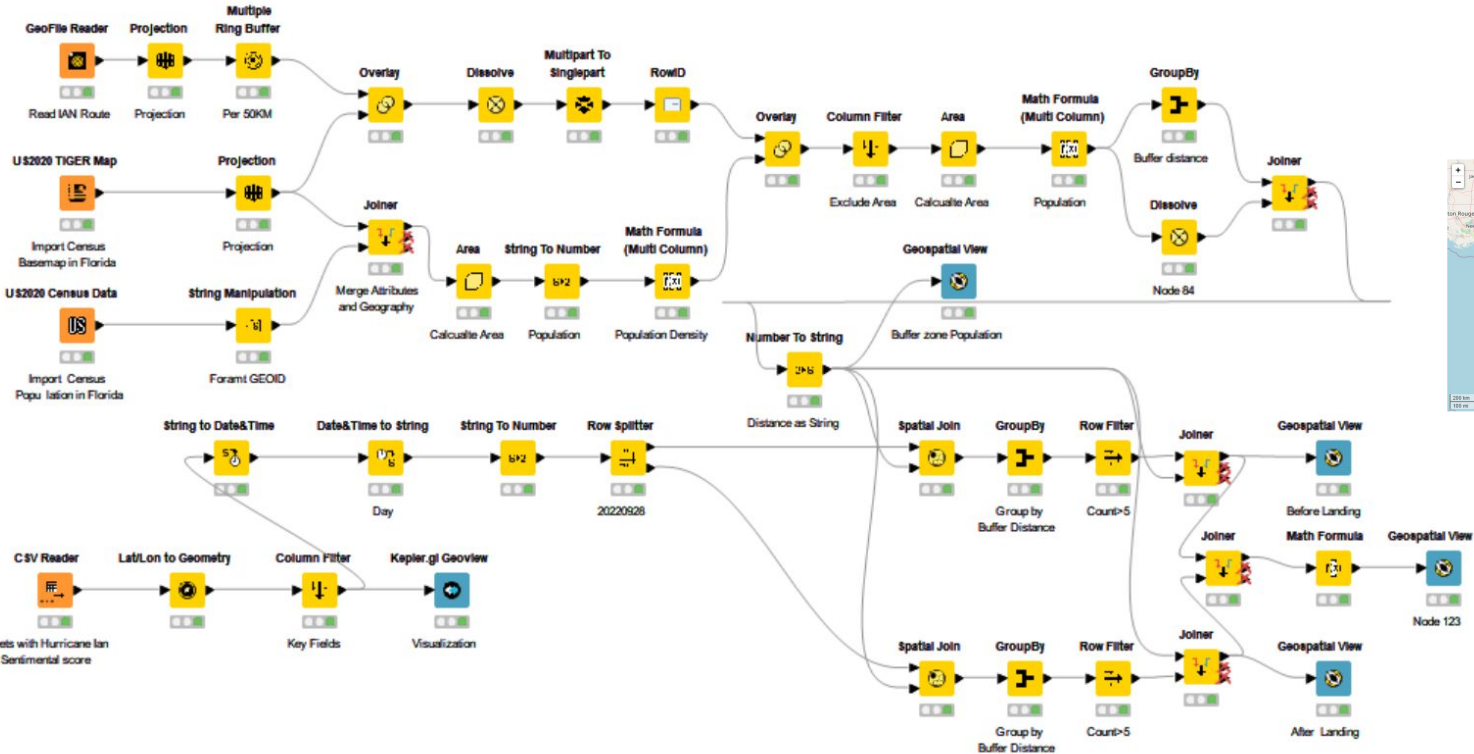
[New Spatial Network Nodes in GA 1.1](#)

<https://www.youtube.com/playlist?list=PLnFUy1f9kH-20dWQGVKkiUA0IbPGxyBUv>



# Impact of Hurricane Ian

## Hurricane IAN- Impact Analysis



# Sponsors and Contributing Open Source Projects



NSF IUCRC  
Spatiotemporal Innovation  
Center(STC)



National Science  
Foundation



Future Data Lab



Open for Innovation

**KNIME**



kepler.gl



GeoPandas



The Python Spatial Analysis Library  
for open source, cross platform  
Geospatial Data Science



# Charting the Future of the SDL

## Sustainable Development:

- Augment the limited team members housed at Harvard
- Search for mechanisms for continued expansion
- Cultivate the communities of development and applications
- Adapt to and evolve with new technology and methodology

## Global collaboration:

- Data collection and service
- Tool development
- Workflow-based case studies development
- Training on spatial data analysis
- Research applications



# The Summer Workshop on Spatiotemporal Innovation

**Date:** Mon - Fri, Jul 10 to Jul 14, 9:00am - 5:00pm

**Location:** 1730 Cambridge Street, Cambridge MA 02138, Cambridge, MA

## Topics (5 days) :

1. An Introduction to Workflow-Based Tools: KNIME Analytics Platform, KNIME Server and KNIME WebPortal
2. Open Data Access, Integration, Online Data Sharing and Visualization
3. Spatiotemporal Analysis with Geospatial Analytics for KNIME
4. Artificial Intelligence Analysis for Spatiotemporal Data
5. Workflow-based Remote Sensing Analysis
6. Business Intelligence Analysis with Spatiotemporal Data
7. Spatiotemporal Analysis with Social Media Data
8. Spatiotemporal Analysis with Healthcare Data
9. Introduction to Affiliate Labs on Spatiotemporal Innovation
10. Discussion on Future Collaborations

[APPLY NOW](#)



**Thank you! Questions?**



# **Workflow Based Tools for Integrated Spatiotemporal Research**

**Wendy Guan, Lingbo Liu & Xiaokang Fu**  
*Center for Geographic Analysis, Harvard University*  
**Shuming Bao**  
*Future Data Lab*

**Acknowledgement:** *This work is partially funded by NSF #1841403*