

2020
International
Smart Cities e-Forum
智慧城市國際線上論壇



Understanding GIS Engineering and Services from Big Data Perspective

Yue, Peng (乐鹏)
Wuhan University

OGC China Forum

←

→

↻

external.opengeospatial.org/twiki_public/ChinaForum/WebHome

LandAdminDWG

MLSdwg

Main

MarineDWG

MassMarket

MetOceanDWG

NREwg

NetCDFu

NordicForum

OGC

PointCloudDWG

QualityOfService

SWE

SWEProCitSci

Sandbox

SmartCitiesDWG

System

TemporalDWG

UKIAP2013

UrbanPlanningDWG

Vocabulary

The OGC China Forum will be open to all OGC member organizations residing in China, or conducting business in China.

Activities

2020

- Environmental Data Retrieval API Standards Working Group (SWG)
 - Github for EDR API: <https://github.com/opengeospatial/Environmental-Data-Retrieval-API>
 - Github for EDR API Sprint (March 18-20): <https://github.com/opengeospatial/EDR-API-Sprint>

2019

- OGC Innovation Program: Disaster Resilience Pilot (DRP-2019)
 - Flood: Emergency Response and Impact Assessment [Demo Slide](#)
 - Youtube channel with the DRP: https://www.youtube.com/playlist?list=PLQsQNjNIDU84zs69bNX__QfZZCTHbq5YN
- Open Standards for Smart Cities: Integrating BIM and Geospatial (GIS), Hong Kong

2018

- OGC China session in the International Workshop on Big Geospatial Data and Data Science (BGDDS 2018), Wuhan

OGC China Forum

← → ↻ ogc.org/ogc/regions/china

Latest membership of the OGC China Forum includes:

- Wuhan University
- Institute of Remote Sensing & Digital Earth, Chinese Academy of Sciences
- Aerospace Information Research Institute, Chinese Academy of Sciences
- Peking University Collaborative Innovation Center for Geospatial Big Data
- Zhengzhou Institute of Surveying & Mapping
- Zhengzhou Zhonghe Jingxuan Information Technology Co. Ltd
- Fuzhou University
- Nanjing Normal University
- Chongqing Survey Institute
- Superman Software Co., Ltd.
- Terra Info Tech (Beijing) Co., Ltd.
- Hong Kong Polytechnic University
- Lands Department, Hong Kong SAR Government
- HKSARG, Development Bureau
- Ribose Inc.

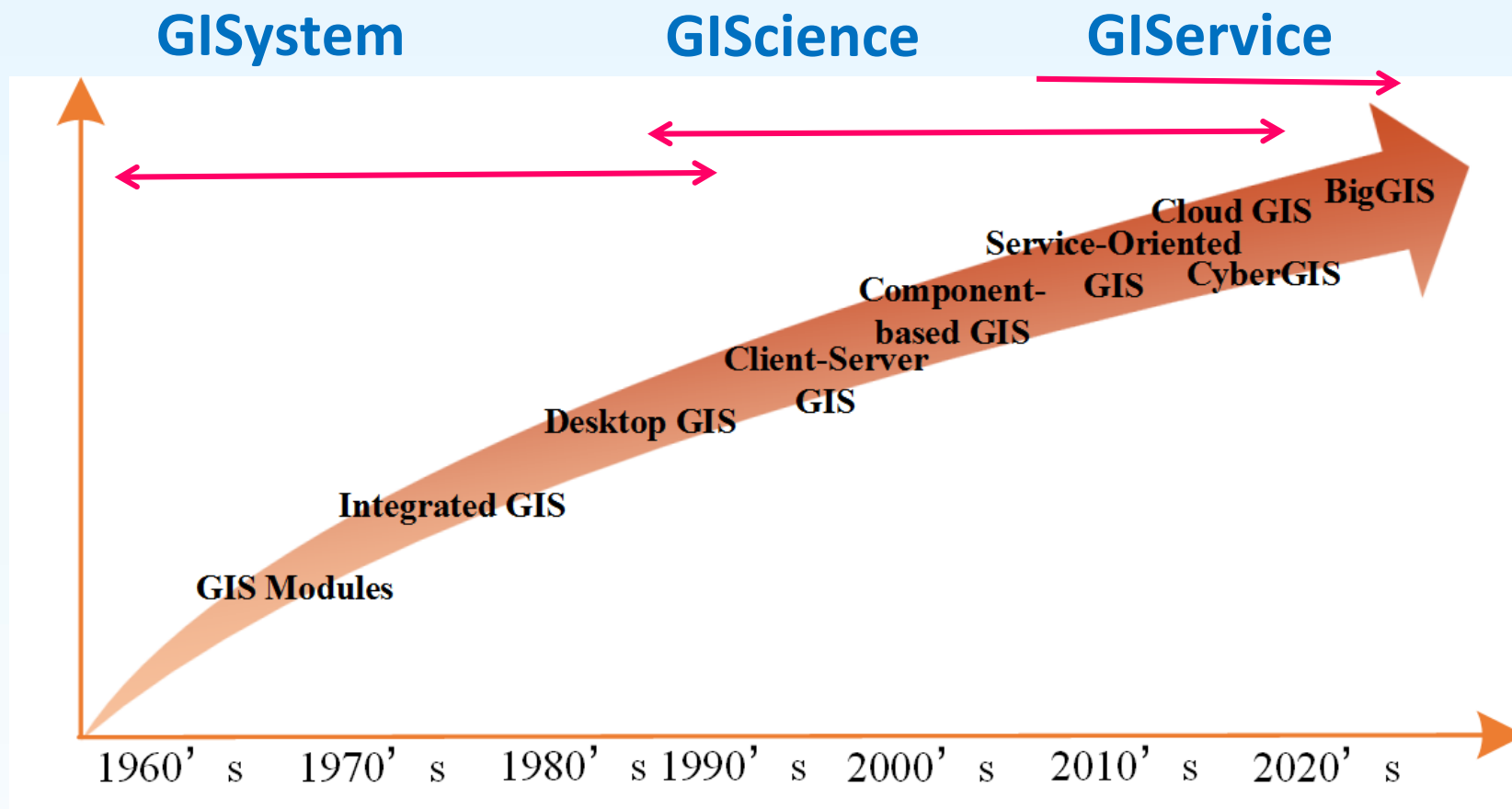
The China Forum is chaired by Peng Yue of Wuhan University. For questions or to become involved, please contact [Peng Yue](#).

The OGC China Forum will be open to all OGC member organizations residing in China, or conducting business in China. For more updates, please visit https://external.opengeospatial.org/twiki_public/bin/view/ChinaForum/WebHome

Content

- 1. Big data infrastructure for GIServices
- 2. Using engineering data for better GIServices
- 3. Cases
- 4. GeoIS platform
- 5. Conclusions

GIS Evolution



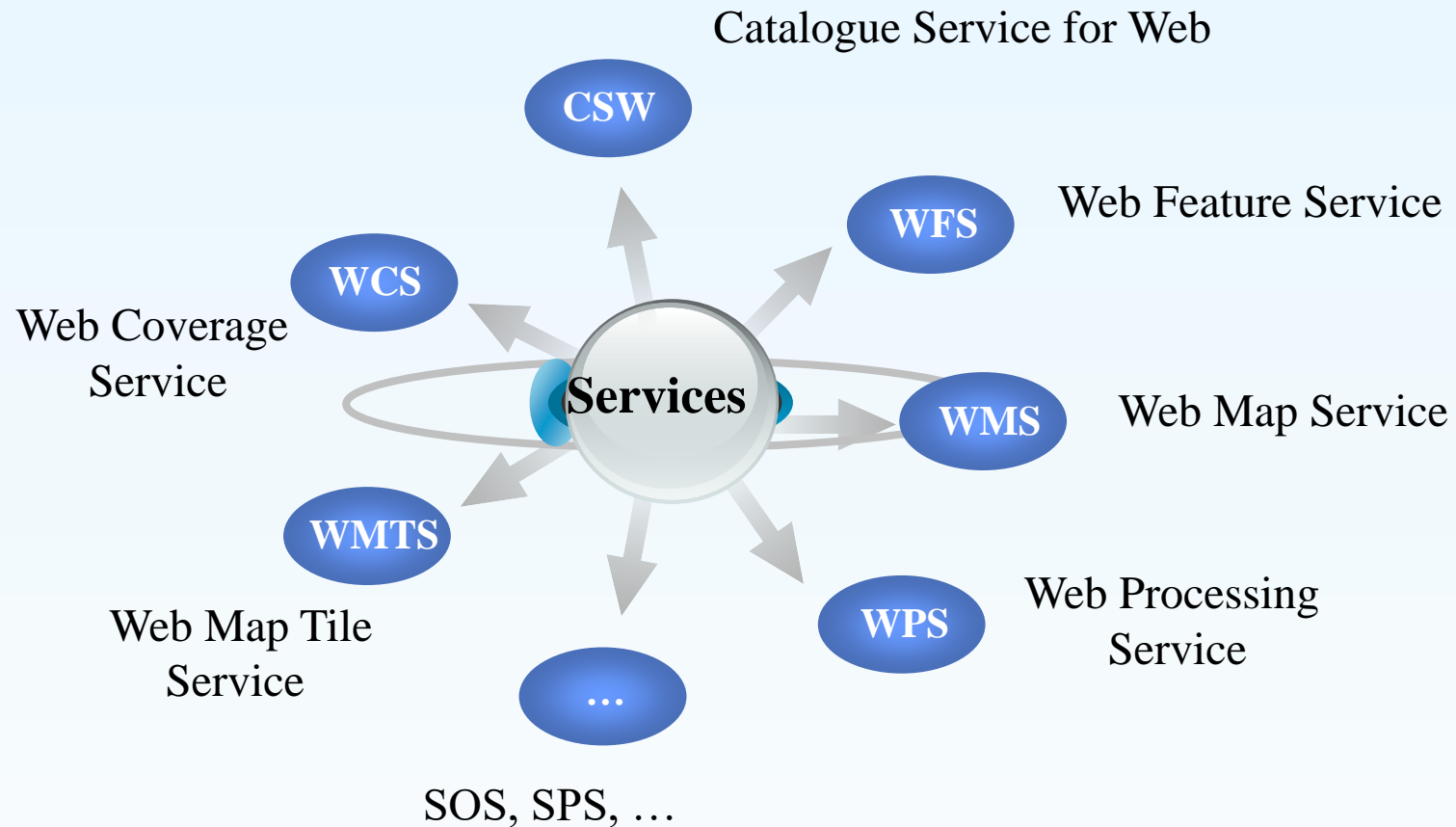
Three Peaks of GIS

GIServices

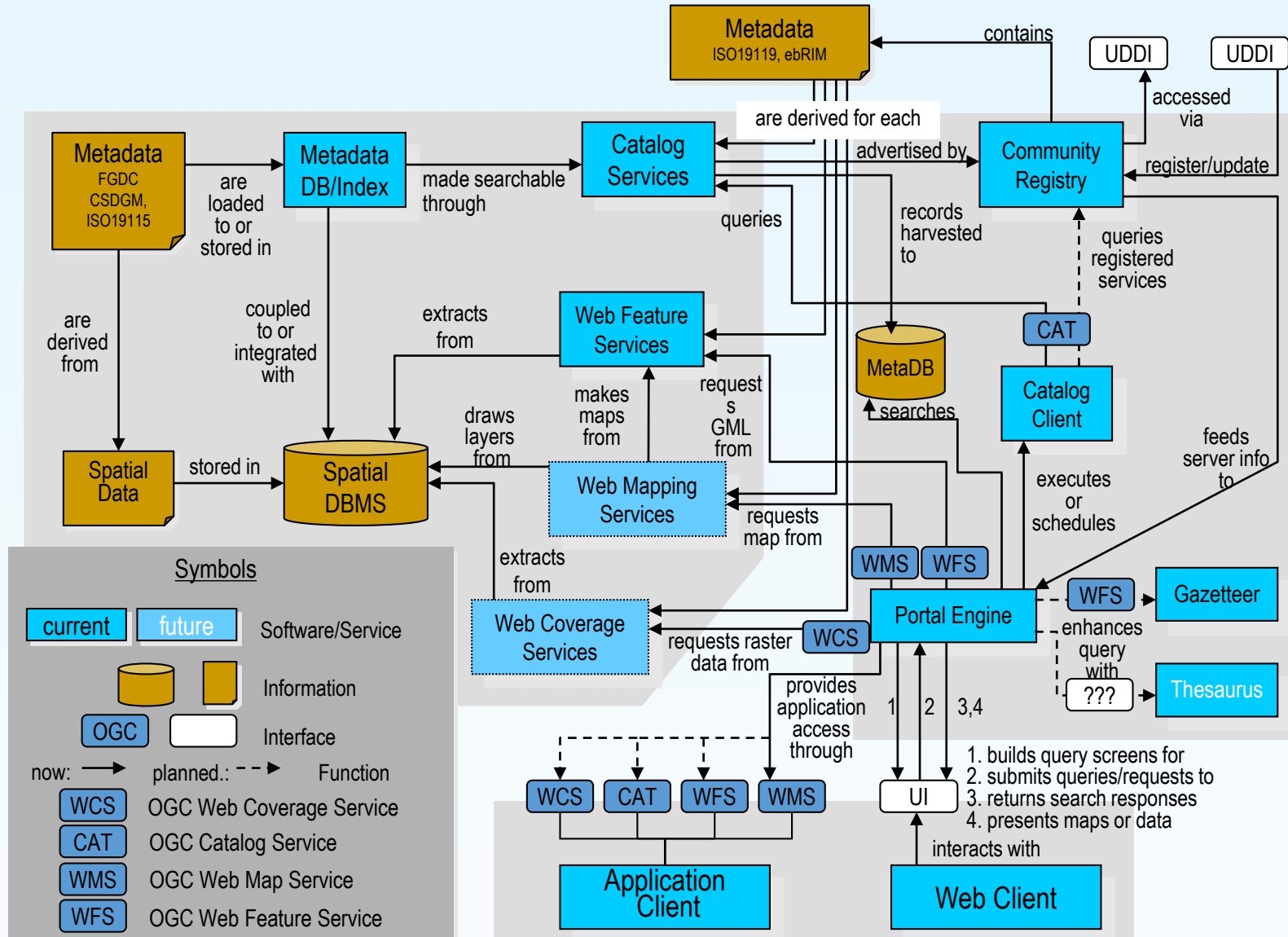
- **Service:** distinct part of the functionality that is provided by an entity through interfaces (ISO)
- **Web service:** software systems designed to support interoperable machine-to-machine interaction over a network (Booth)
- **GIService:** functional software entities with interfaces that can provide geospatial data, information, and knowledge across the Web

Traditional GIServices

Interfaces/Interoperability are important



SDI (Spatial Data Infrastructure) Layered on Services

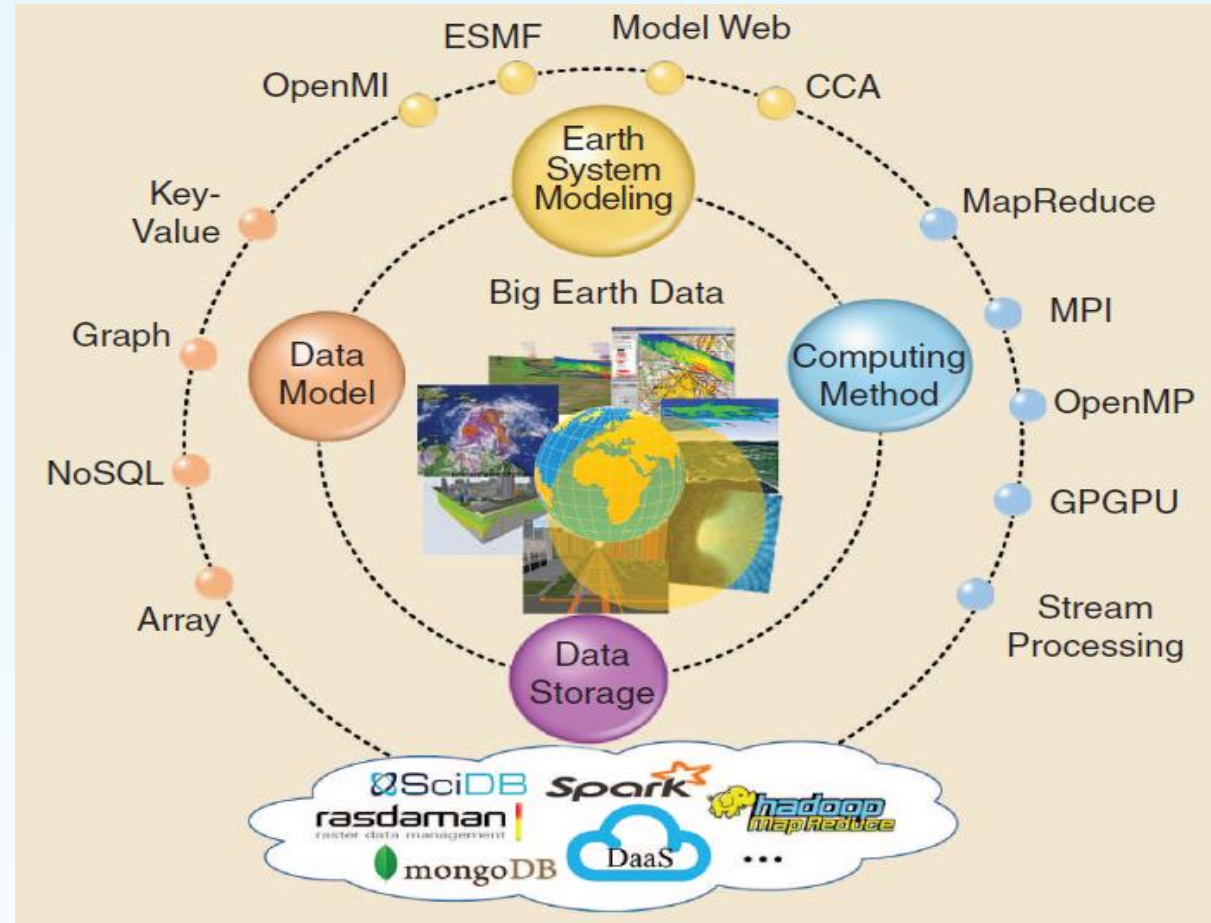


Big Data Context

From “SDI above services” to “BDI behind services”

Big Data Infrastructures(BDI)

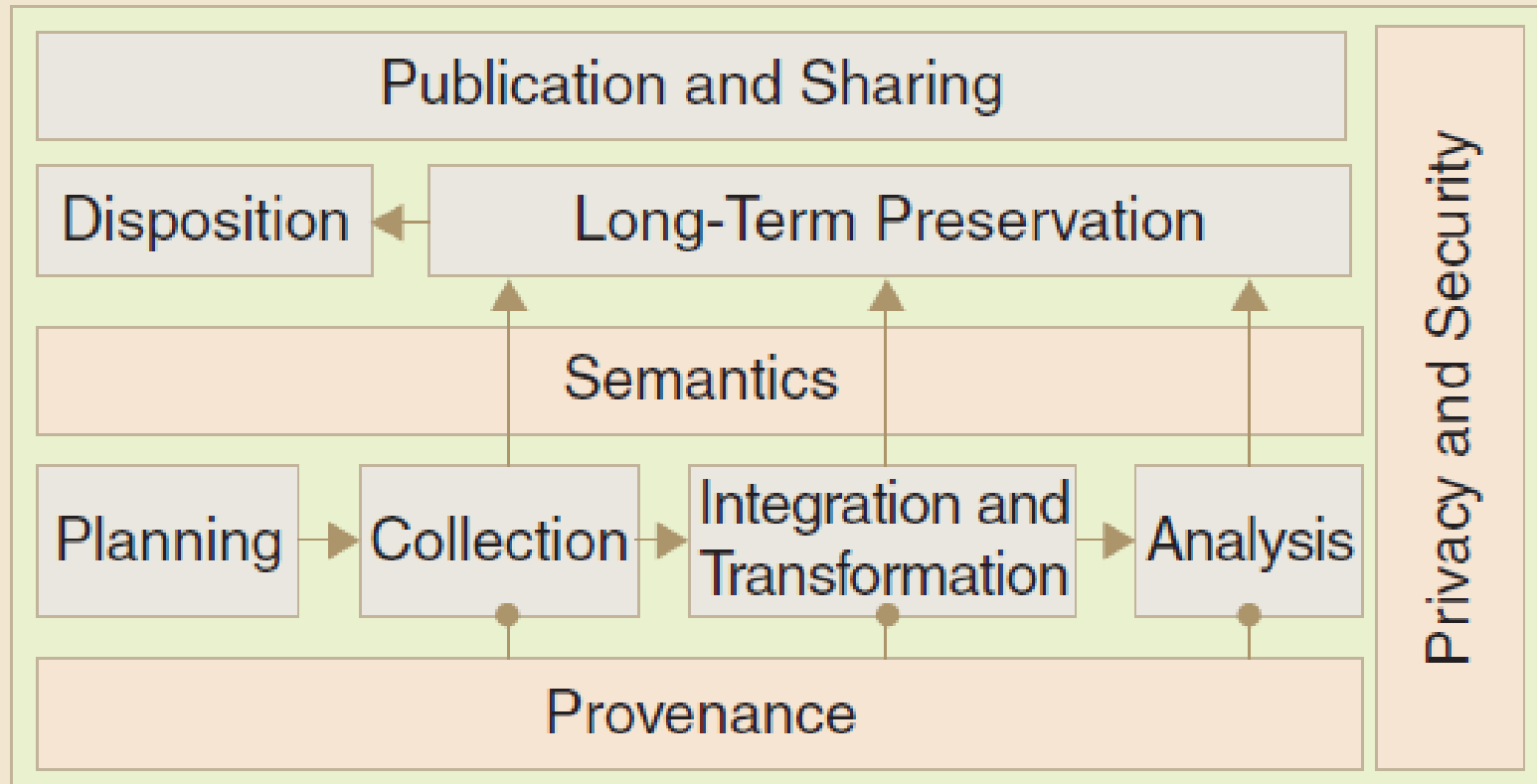
- **Data infrastructure**
- **Computing infrastructure**
- **Modelling infrastructure**



Recent Activities in Earth Data Science. Yue et al. 2016. IEEE GRSM, 4(4)

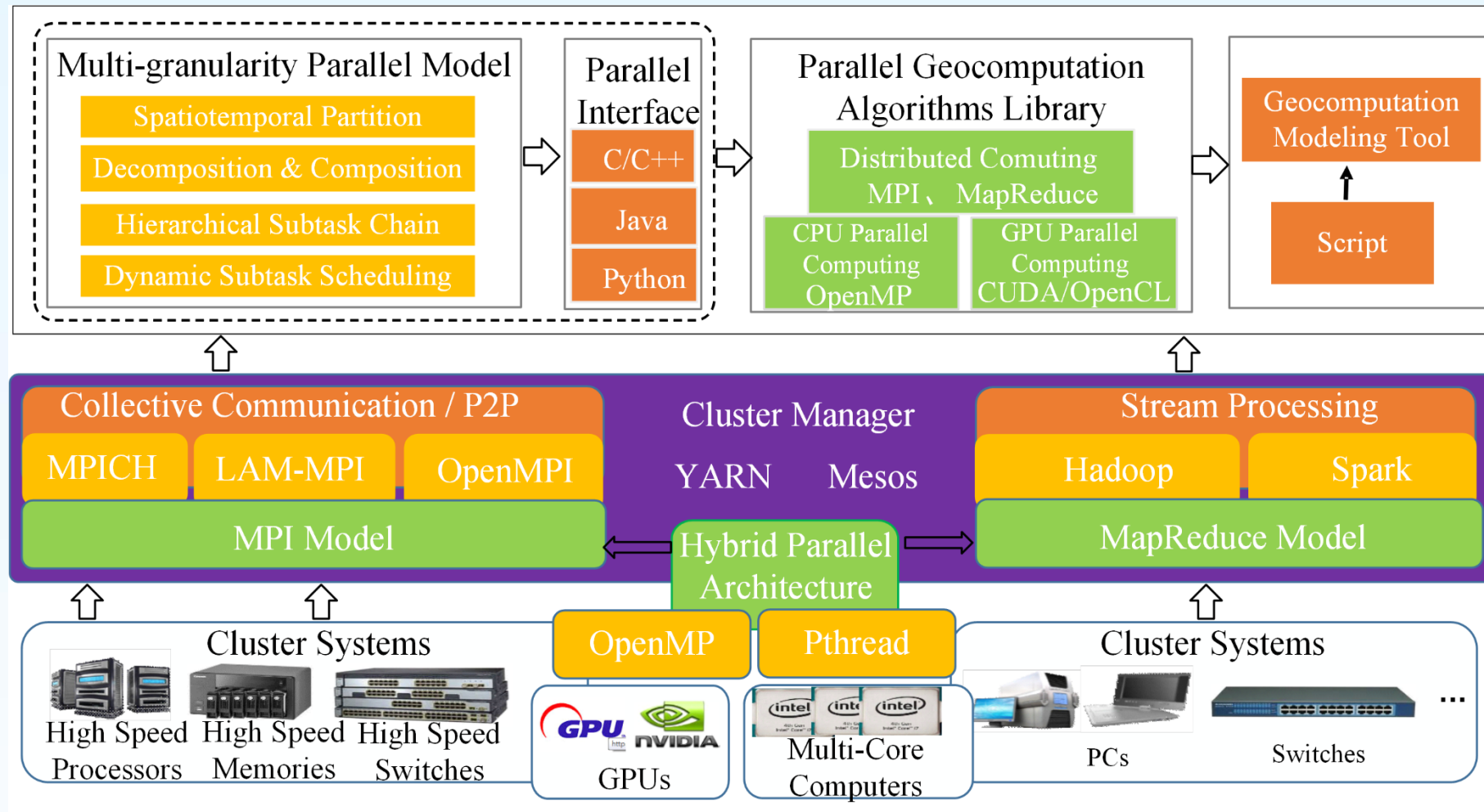
Big Data Context

Infrastructures dealing with data life cycle



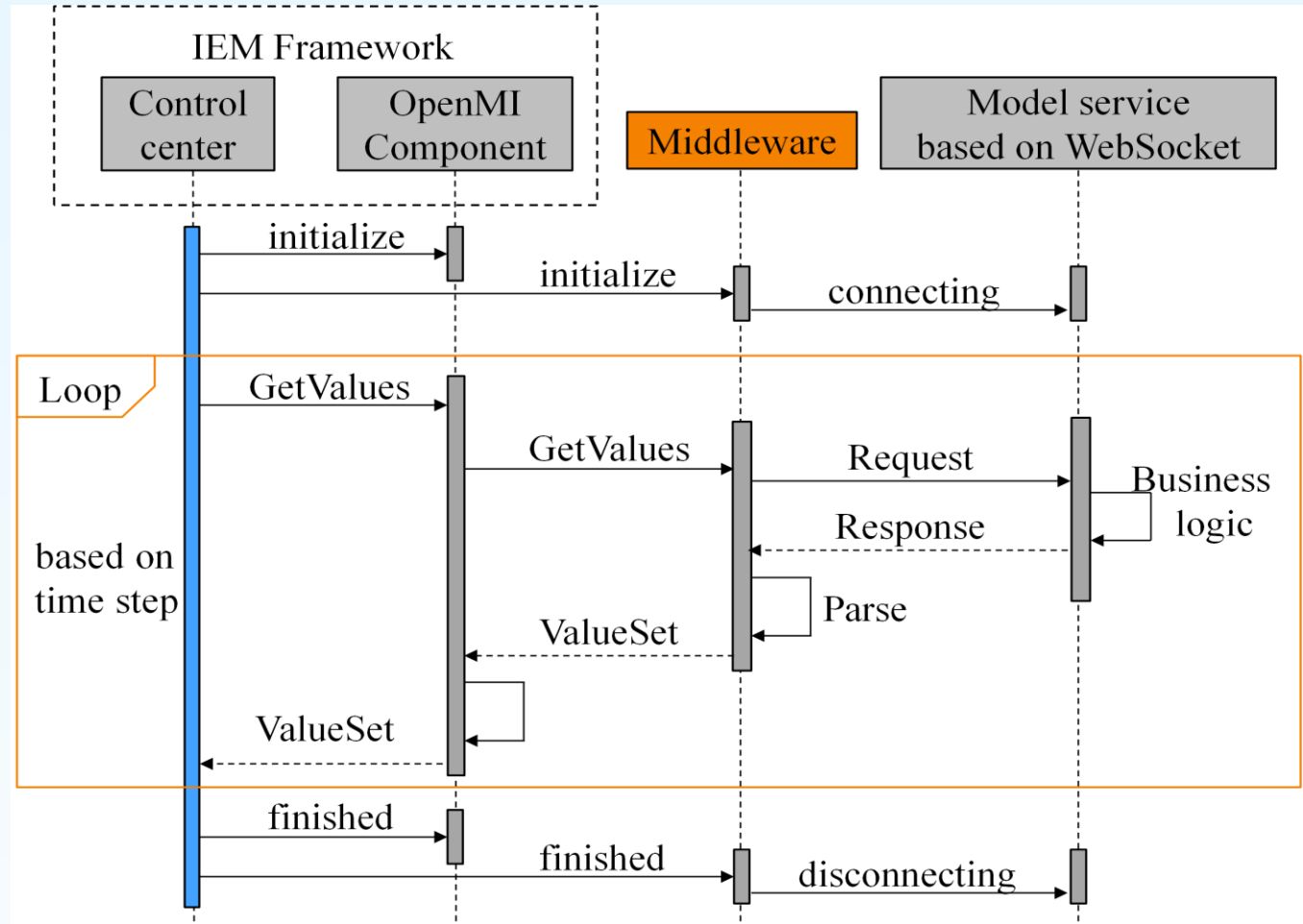
Big Data Context

Infrastructures dealing with big data analytics



Big Data Context

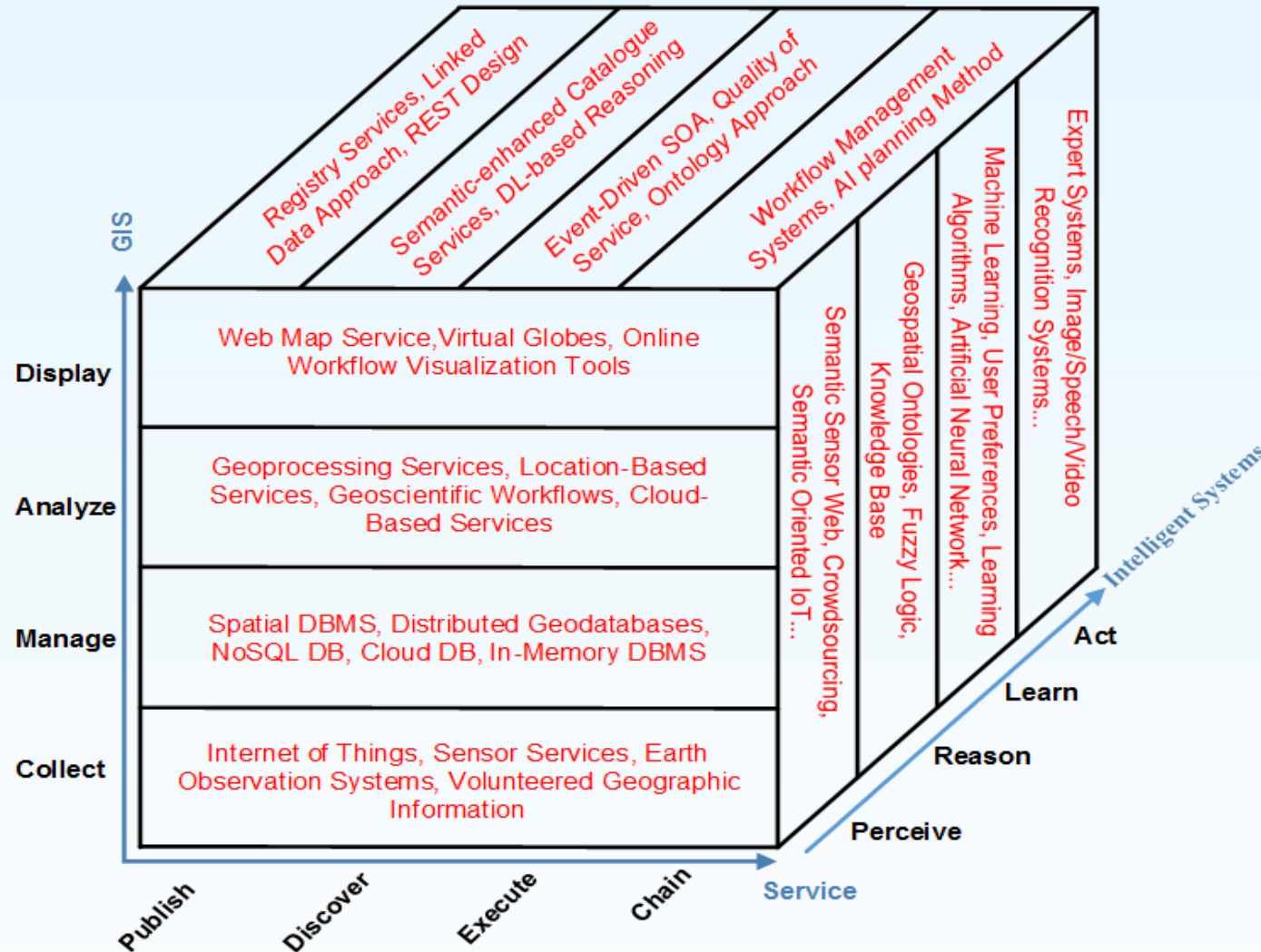
Infrastructures dealing with Model Web



Coupling components and services for integrated environmental modelling. Gao et al. 2019. ENVSOFT, 118

Big Data Context

Hybrid infrastructures support Intelligent GIServices



Towards Intelligent GIServices.
Yue et al. 2015. ESIN, 8(3)

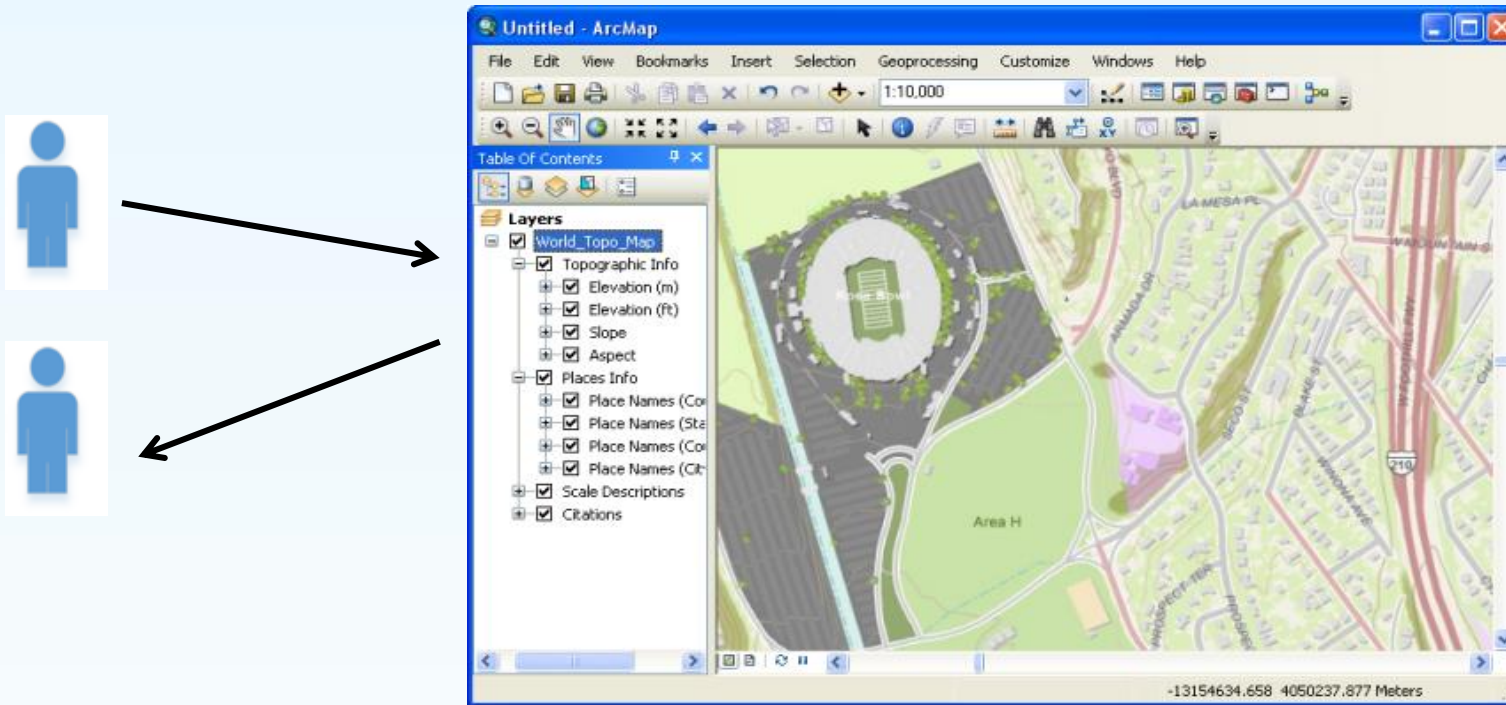
GIS Engineering

A little bit more engineering for GIS

- **GIS engineering combines GISystem, GIScience, and GIServices**
- **GIS engineering has become a bachelor degree program in China in 2016, due to the increasing market for GIS engineers in many engineering projects**
- **GIS engineering is more important compared to its development decades ago due to wider applications**

Traditional GIS Engineering

- Software engineering activities
- Hundreds of thousands GIS projects developed



Engineering is yet another big Data

Big Data Context

New questions:

- **Can we get the GIS engineering data?**
- **How much is reused from past GIS projects?**
data, algorithms, workflows, GUIs, infrastructures...
- **How can we expedite development of GIS projects?**
cost, schedule, mutual enrichment between GeoAI and GIS engineering,
concept drift, metamorphic testing...
- **Can we recommend best practices?**
Imbalance between published algorithms versus workable ones in
projects...

Big Data Context

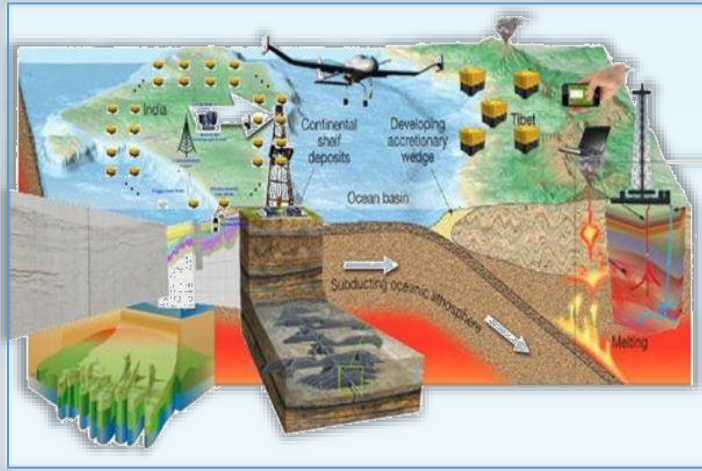
Two bullets:

- Primarily we need make big GIS engineering data sensible and accessible
- An infrastructure for GIS engineering is needed

A little engineering data goes a long way



Cases: natural disasters



A **natural disaster** is a major adverse event resulting from natural processes of the Earth; examples are floods, hurricanes, tornadoes, volcanic eruptions, earthquakes, tsunamis, and other geologic processes. A natural disaster can cause loss of life or property damage, and typically leave some economic damage. [NOAA, USA]

- **Complex process and mechanism**
- **Wide coverage, large mount of secondary disasters**
- **Real-time timeliness requirement of disaster monitoring and response**



Earthquake



Flood

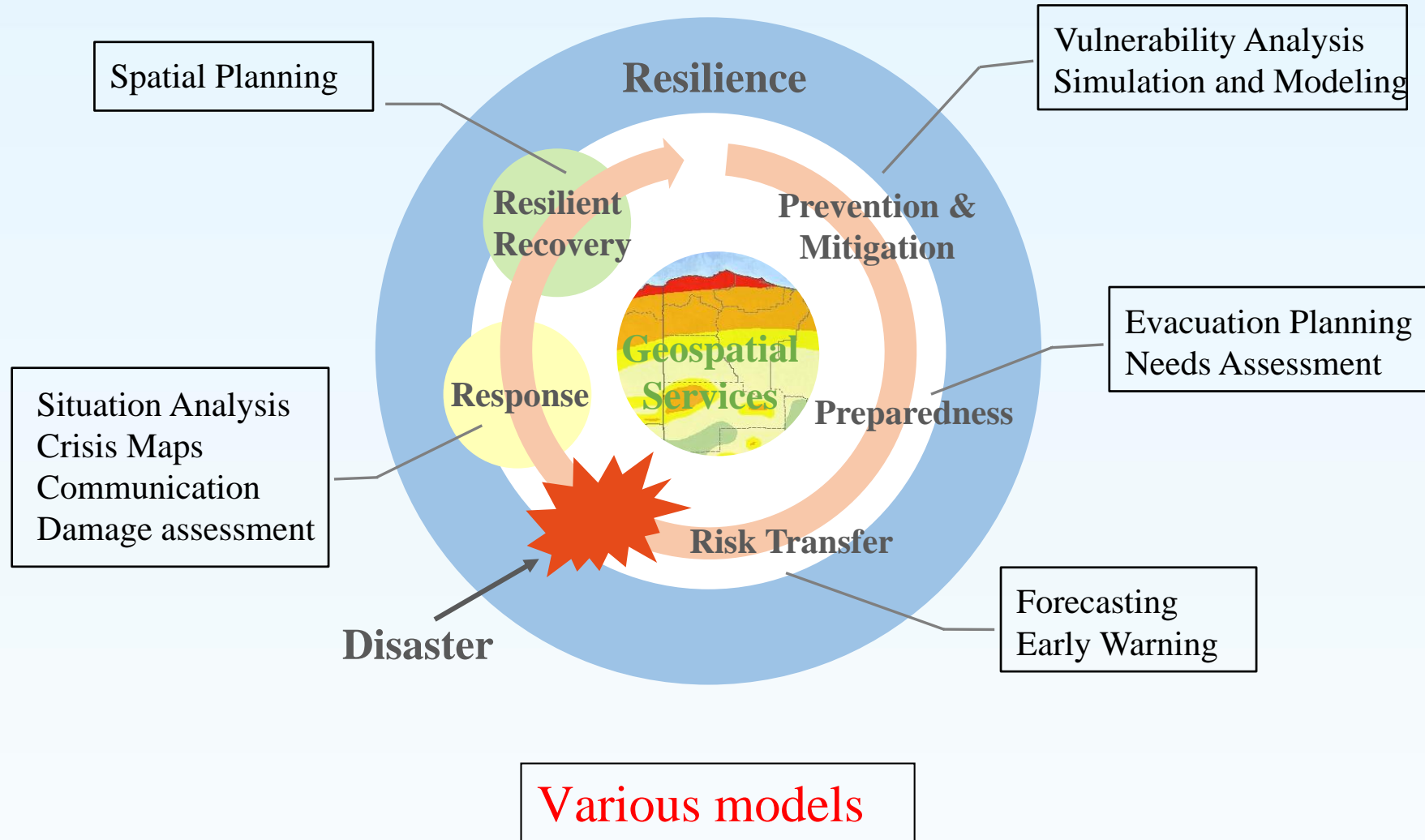


Landslide



Barrier lake

Cases: natural disaster risk management

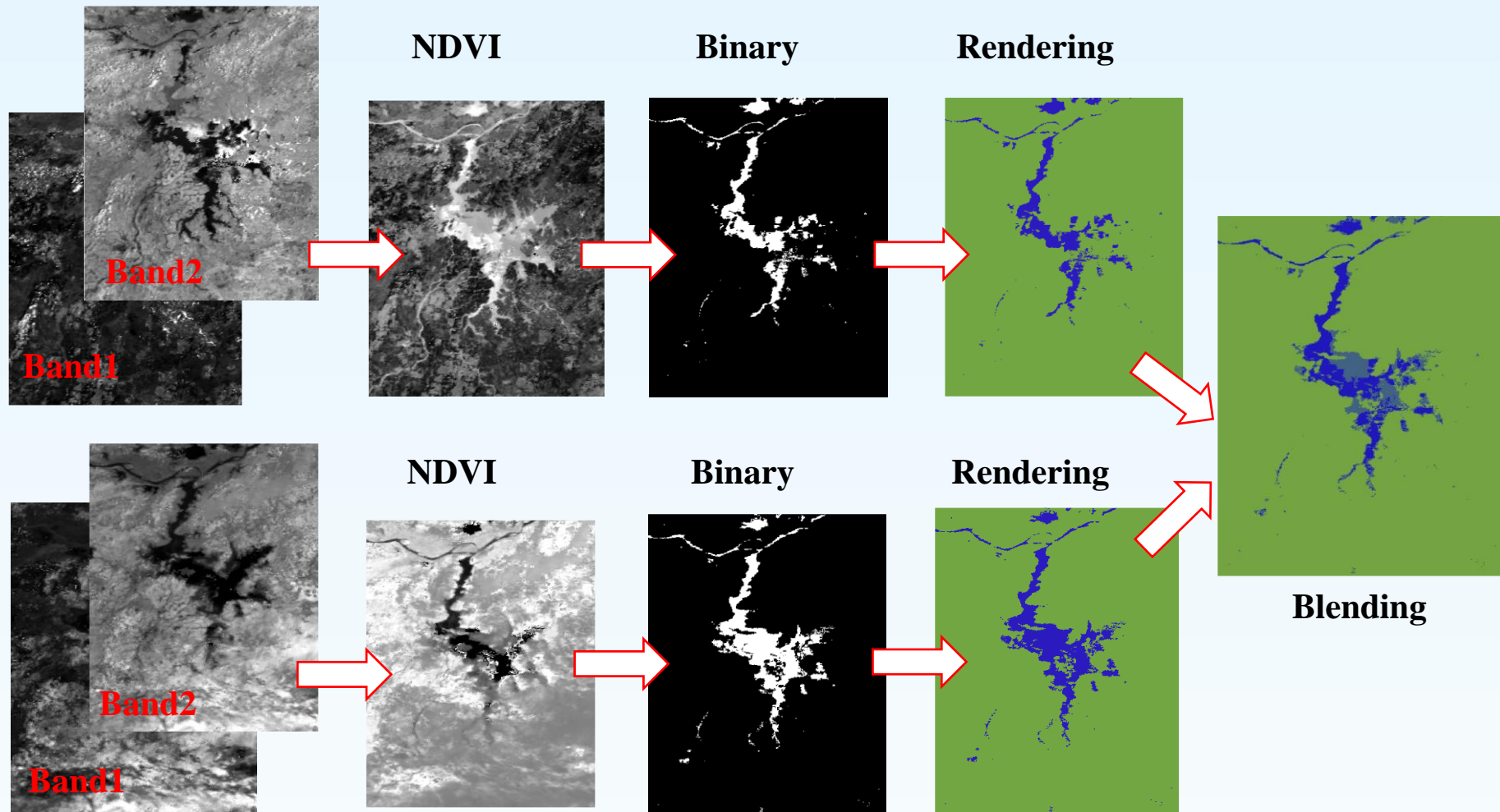


One Belt One Road Disasters

- According to the international disasters database EM-DAT, **4580 large disasters** occurred in the One Belt One Road region from 1980 to 2015. **Floods, storms, and earthquakes** have top numbers, representing respectively **37.28%, 24.86% and 12.81%** of the total number of events.

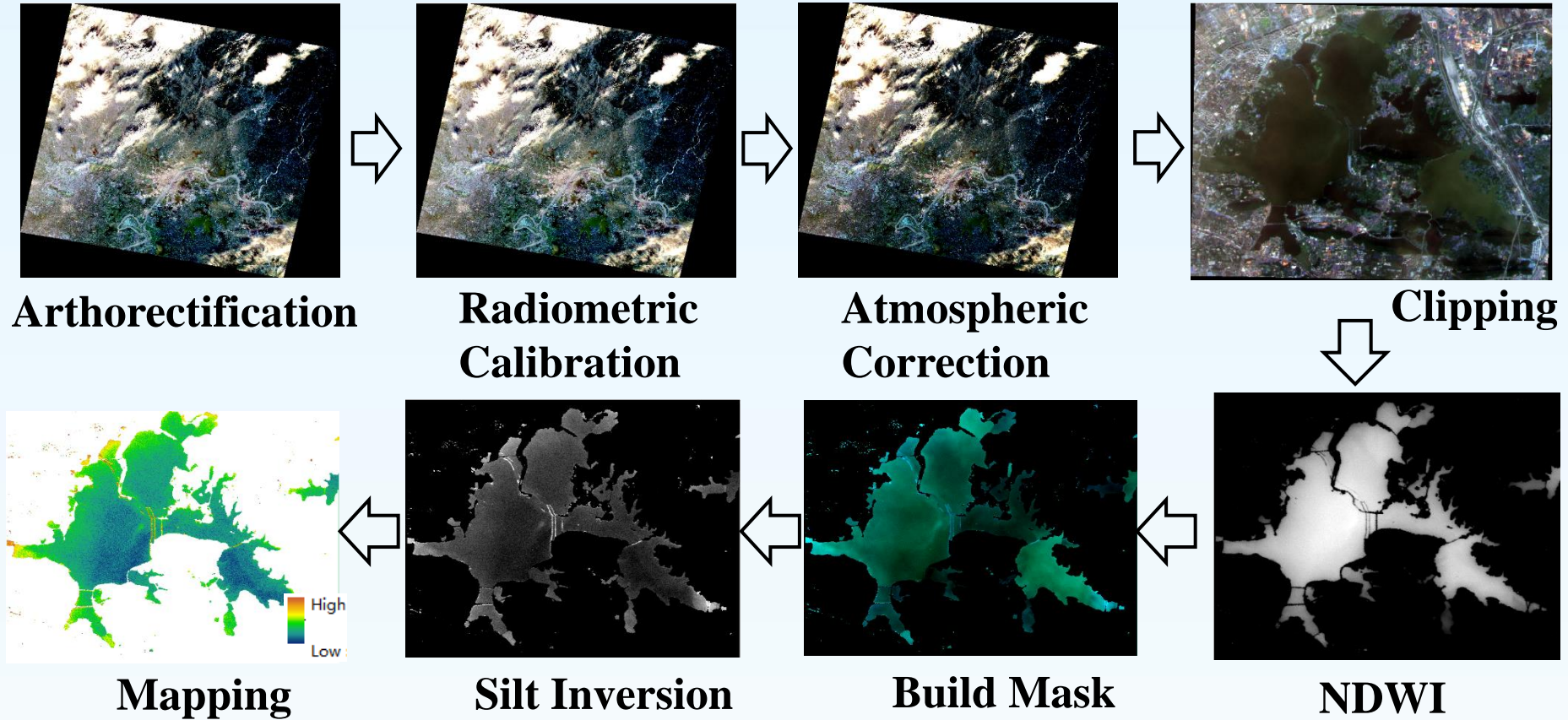


Various Processing Flows



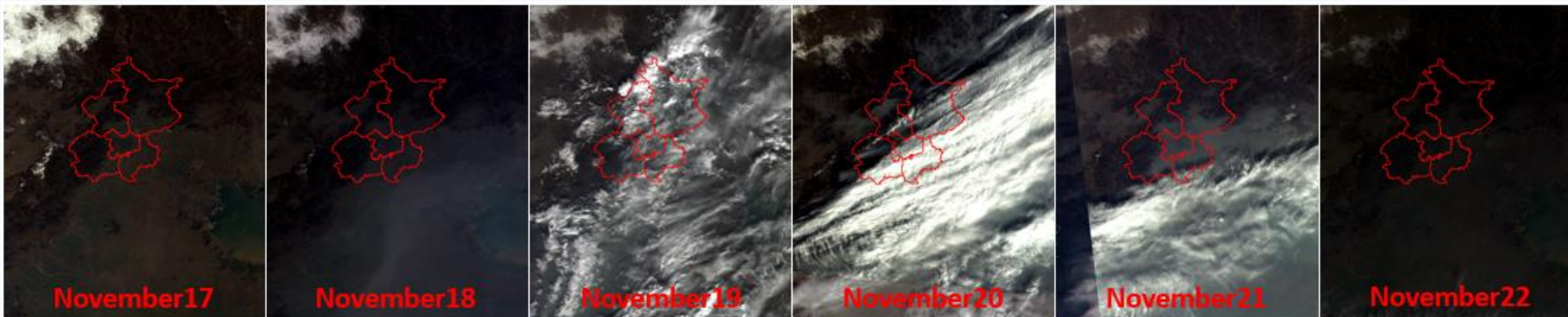
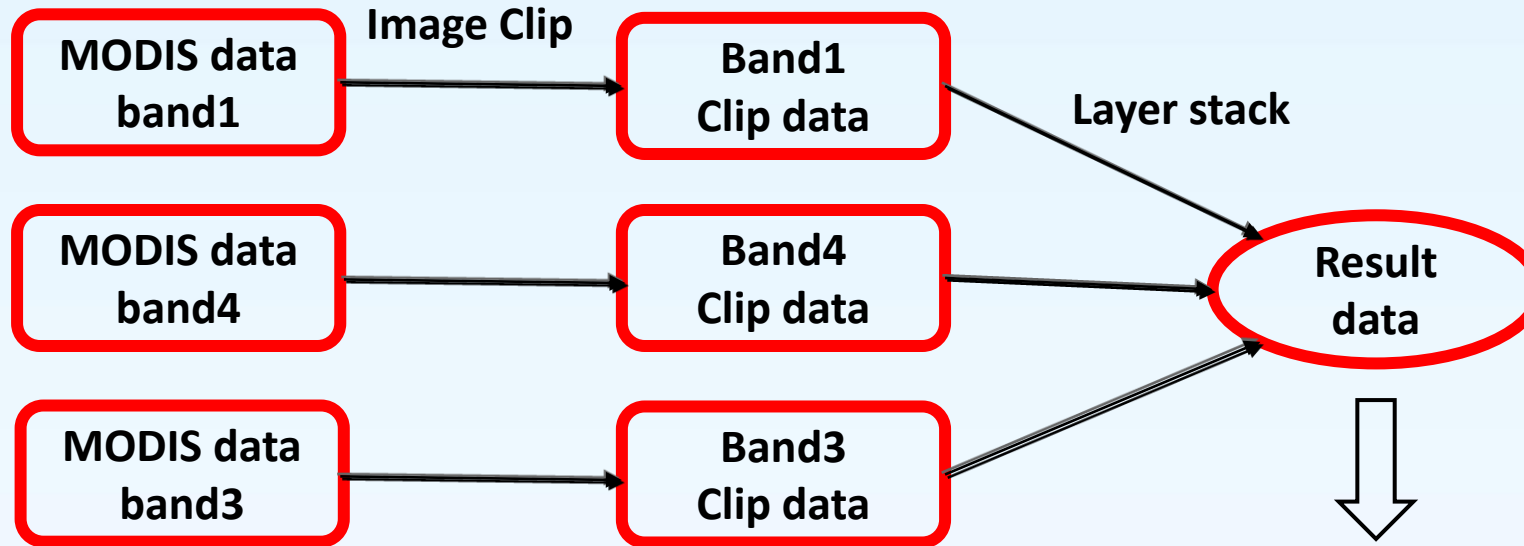
A workflow for change extraction of water coverage area from remote sensing imagery

Various Processing Flows



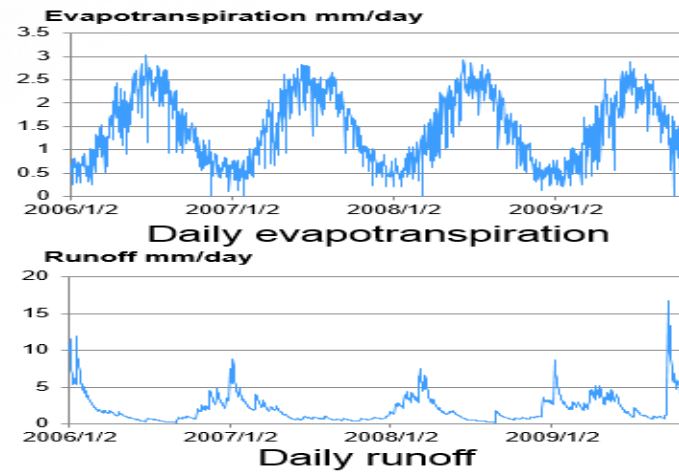
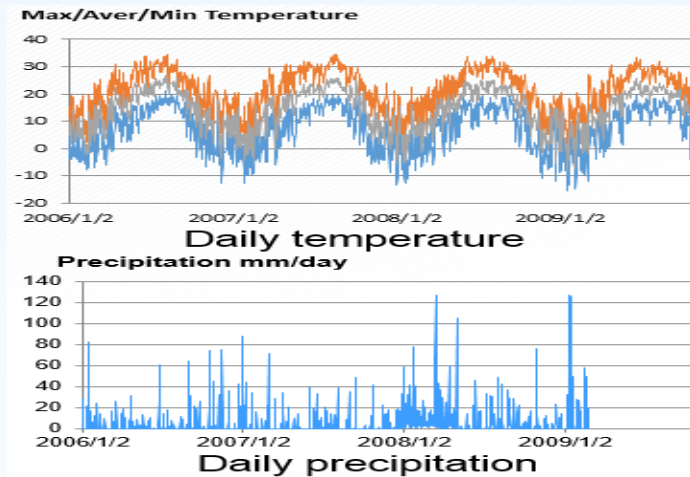
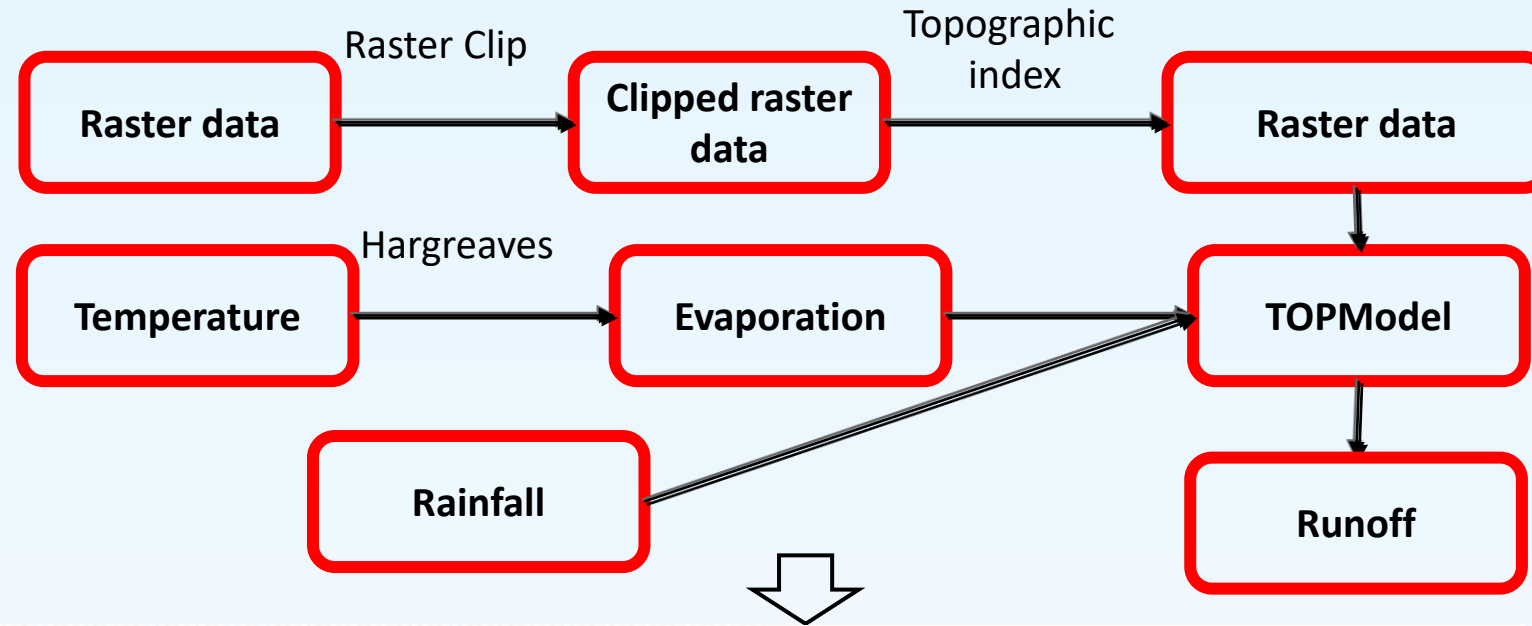
Turbidity extraction from the remote sensing imagery

Various Processing Flows



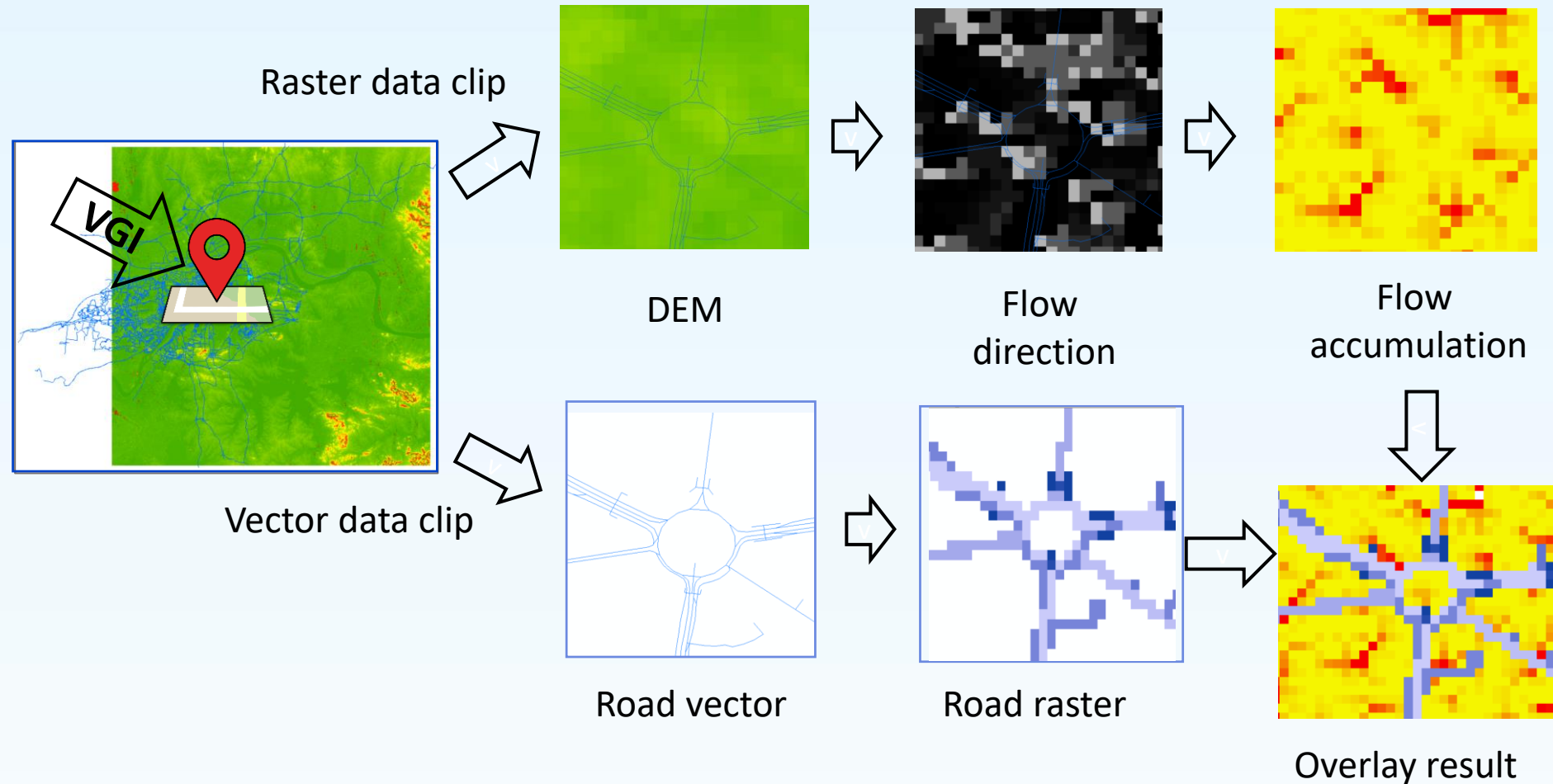
A workflow for haze detection from remote sensing imagery

Various Processing Flows



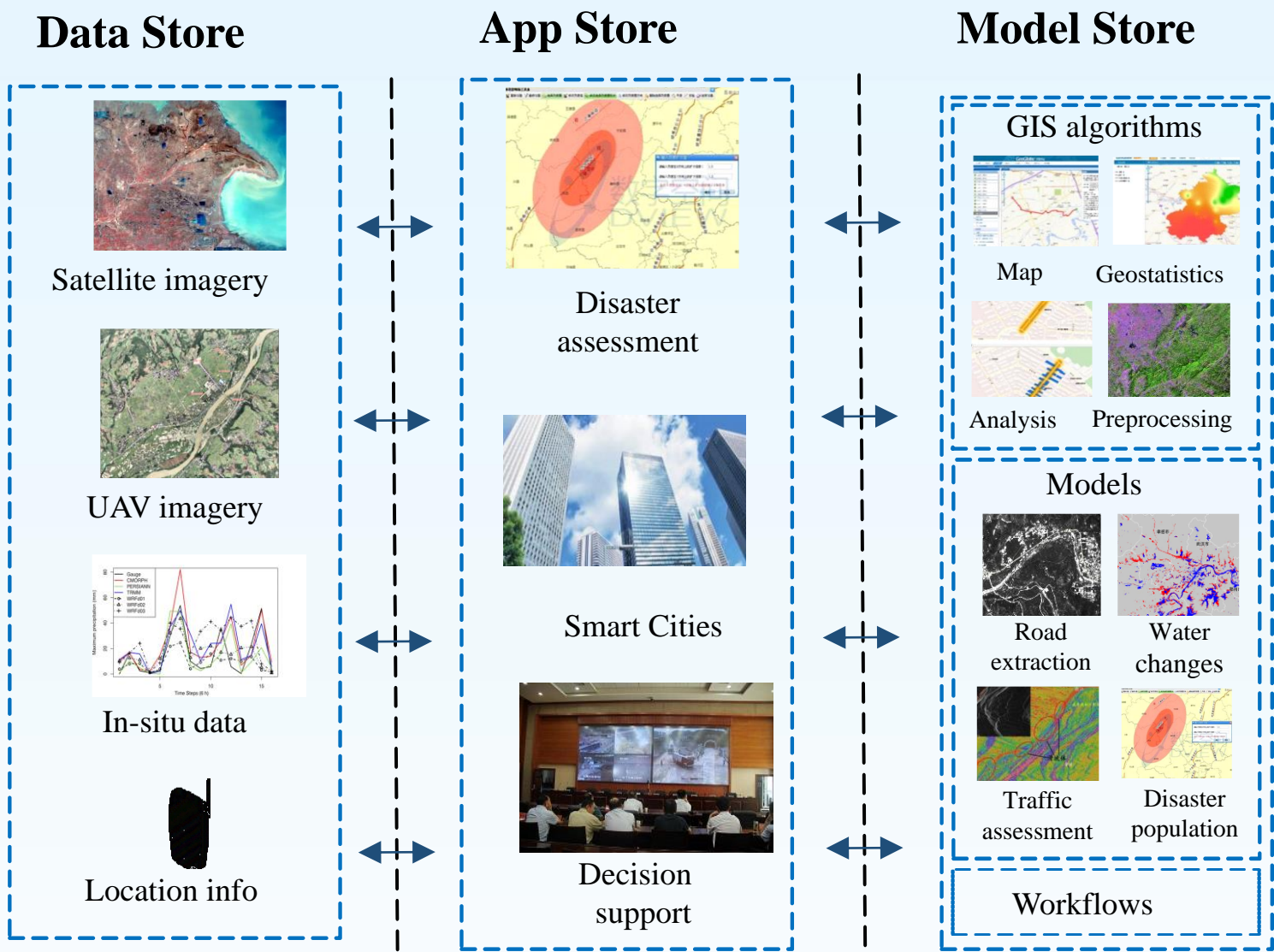
A workflow for watershed runoff simulation

Various Processing Flows

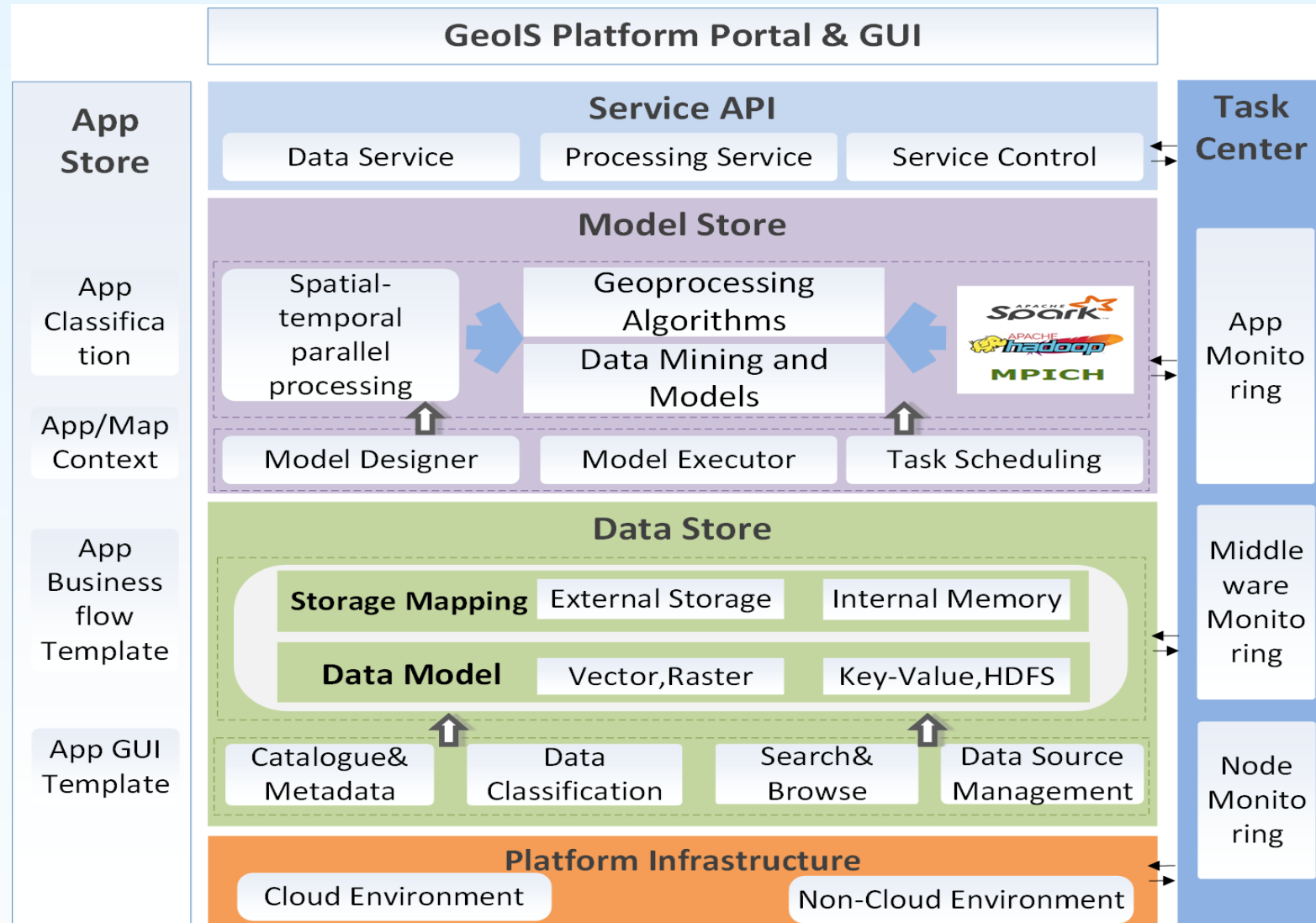


A workflow for urban waterlogging from DEM and vector data

Creating stores holding GIS engineering data



GeoIS: Geospatial big data enabled Intelligent Service platform

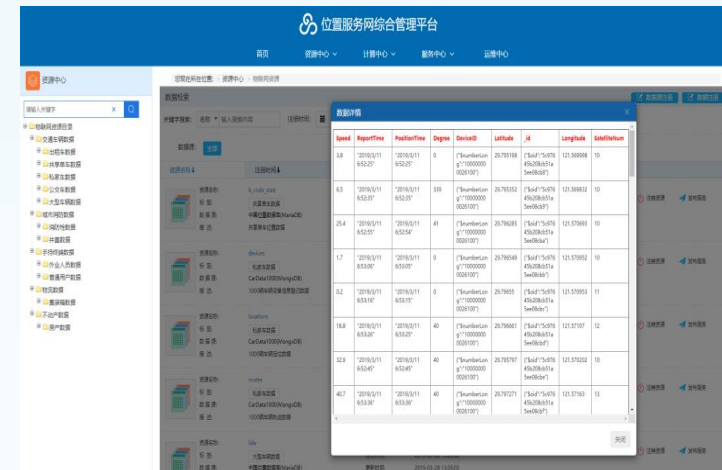


GeolS

Data Store :

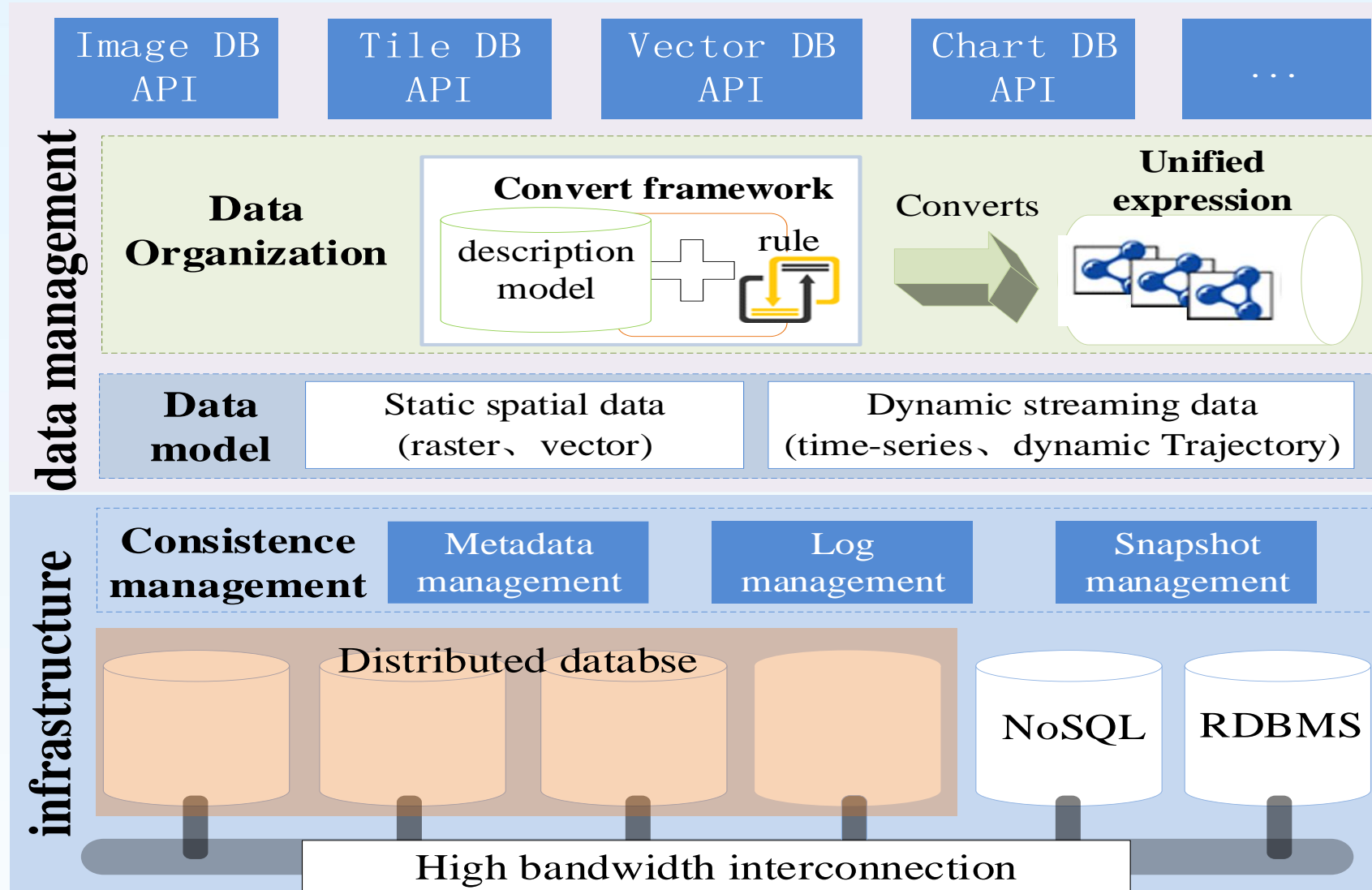
Accommodating various geospatial data resources, heterogeneous data, and real-time observations, using distributed nodes

- Metadata-based Warehouse management
- Large distributed geospatial data storage: In-memory SpatialRDD, External storage HDFS/HBase
- Tag based topic management
- Search: Category search, full text search
- Browse: data browse, map browse
- More data source plug-ins: Oracle, ArcSDE , MySQL , MongoDB, PostgreSQL, MariaDB, Netdisk file, OGC data/sensor services

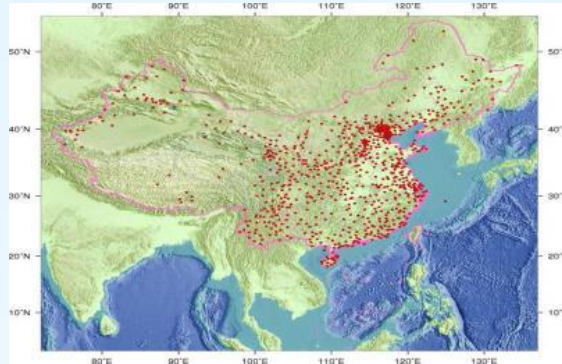


GeoS

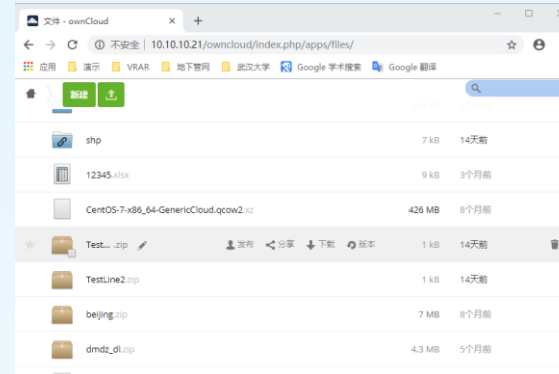
Distributed Data Storage



GeolS Data Publication



**Coordinated
monitoring**



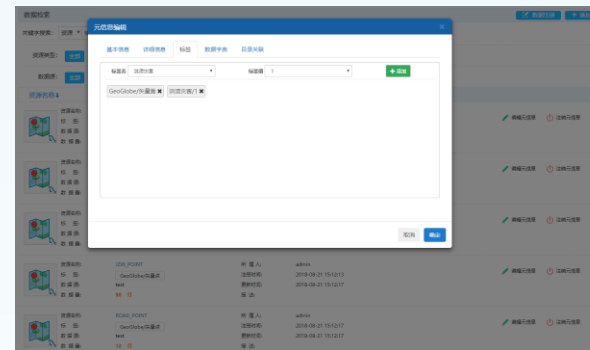
**Upload data to
Network disk**



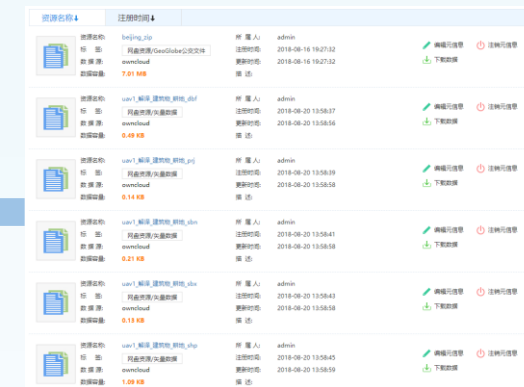
**Extract-Transform-Load
(ETL)**



Web service



Data annotation



Data classification

Model Store :

Libraries of algorithms, models, tools, workflows, and their high performance implementation in computational infrastructures

The screenshot displays the GeolS Model Store interface, which is a web-based platform for managing and executing geospatial services. The interface is divided into several sections:

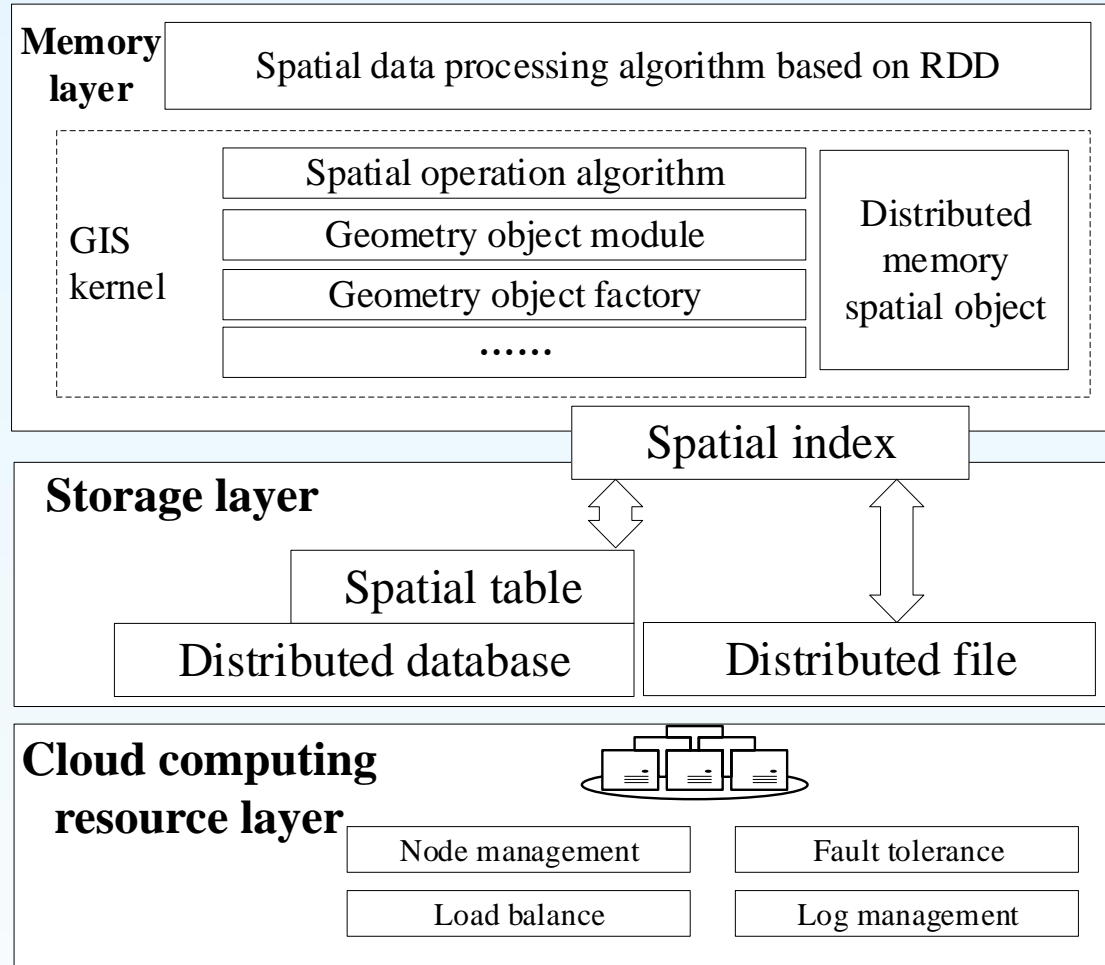
- Header:** The top navigation bar includes the GeolS logo and the title "位置服务网综合管理平台" (Geospatial Service Network Comprehensive Management Platform). It also contains links to "首页" (Home), "资源中心" (Resource Center), "计算中心" (Computation Center), "服务中心" (Service Center), and "运维中心" (Operation Center).
- Left Sidebar:** A sidebar titled "服务中心" (Service Center) lists various service categories, including "处理服务目录" (Processing Service Directory), "主题提取服务" (Theme Extraction Service), "数据关联服务" (Data Association Service), "相似度计算模块服务" (Similarity Calculation Module Service), "字符相似度计算服务" (Character Similarity Calculation Service), "位置相似度计算服务" (Location Similarity Calculation Service), "空间相似度计算服务" (Spatial Similarity Calculation Service), "时间相似度计算服务" (Temporal Similarity Calculation Service), "数据转换服务" (Data Conversion Service), "信息融合服务" (Information Fusion Service), "结构化语义检索服务" (Structured Semantic Retrieval Service), "多传感器数据融合服务" (Multi-sensor Data Fusion Service), "时空变化趋势分析服务" (Spatiotemporal Change Trend Analysis Service), "动态可视化服务" (Dynamic Visualization Service), "基础空间处理服务" (Basic Spatial Processing Service), "空间数据编码服务" (Spatial Data Encoding Service), and "北斗编码服务" (Beidou Encoding Service).
- Main Content Area:** The main area displays a list of services. A modal window titled "服务详情" (Service Details) is open, showing the following information:
 - 服务地址:** http://202.114.118.190:18080/api/service/model_resource/74
 - 访问类型:** POST
 - 输入参数:**

参数名称	参数类型	参数说明
latitude	Double	坐标点纬度
longitude	Double	坐标点经度
height	Double	坐标点高程
codesize	Integer	编码长度
 - 输出参数:**

参数名称	参数类型	参数说明
code	String	坐标点的三维北斗编码

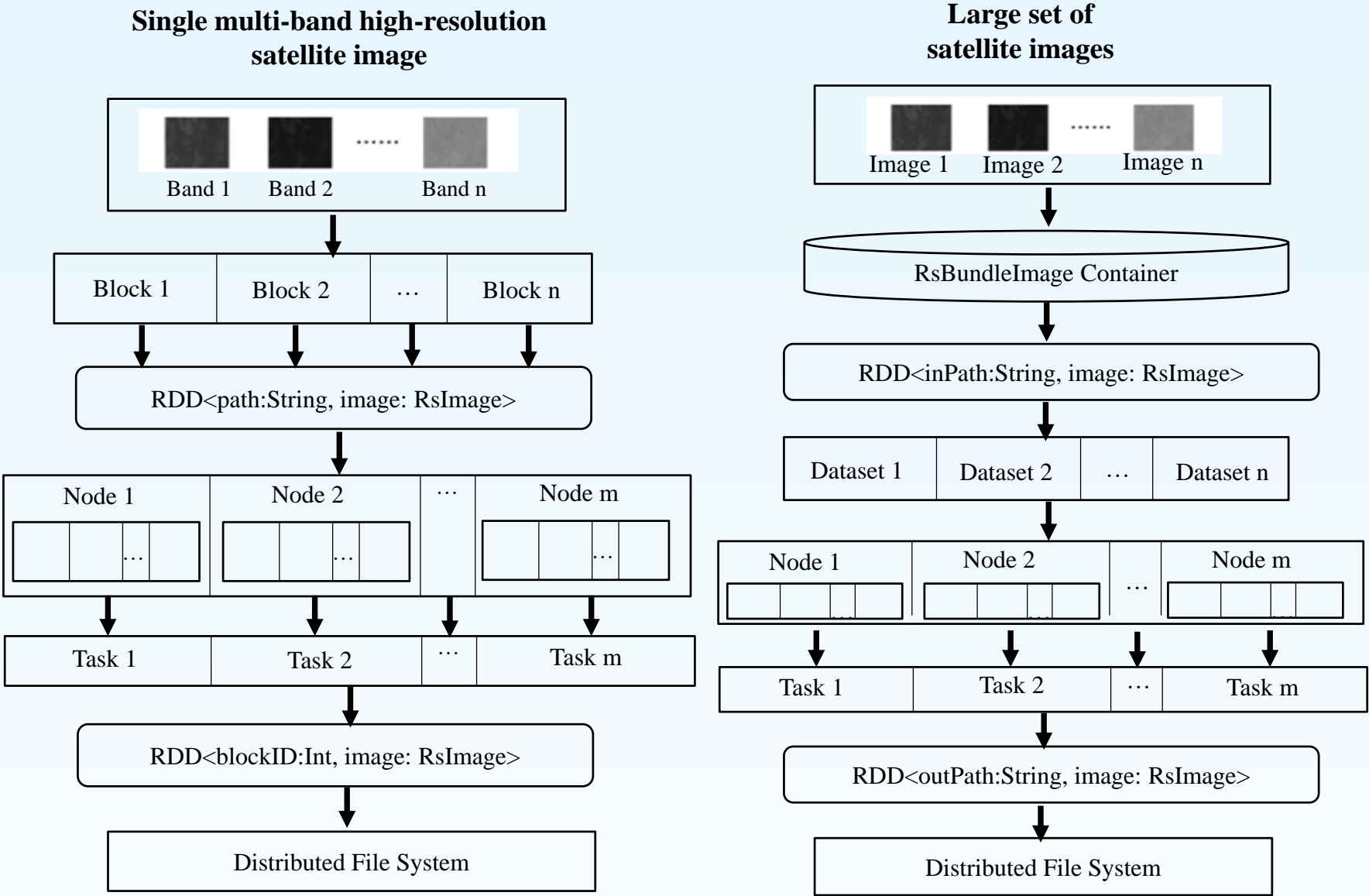
- Right Panel:** The right panel shows a workflow diagram titled "计算中心" (Computation Center). It displays a sequence of steps, including "输入关键字" (Input Keyword), "分析应用模型" (Analyze Application Model), "提取特征" (Extract Features), "位置相似度计算" (Location Similarity Calculation), "空间相似度计算" (Spatial Similarity Calculation), "时间相似度计算" (Temporal Similarity Calculation), "数据转换" (Data Conversion), "信息融合" (Information Fusion), "结构化语义检索" (Structured Semantic Retrieval), "多传感器数据融合" (Multi-sensor Data Fusion), "时空变化趋势分析" (Spatiotemporal Change Trend Analysis), "动态可视化" (Dynamic Visualization), "基础空间处理" (Basic Spatial Processing), "空间数据编码" (Spatial Data Encoding), and "北斗编码" (Beidou Encoding). The diagram shows how these steps are interconnected to form a complete workflow.

GeoIS: SPARK Processing for Vector Data

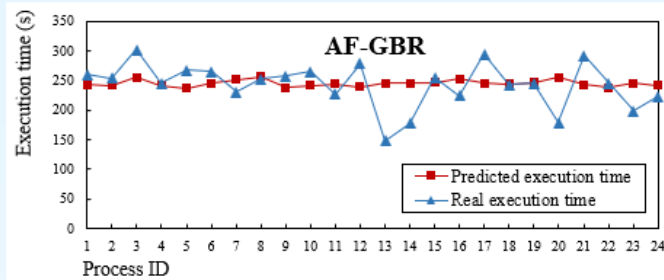


- **Storage vector data in HDFS and HBase with self-adapted GeoHash indexes**
- **Mapping vector to distributed memory spatial objects with Spark RDD**
- **Integrating GIS kernels with designed RDD to achieve distributed spatial computing**

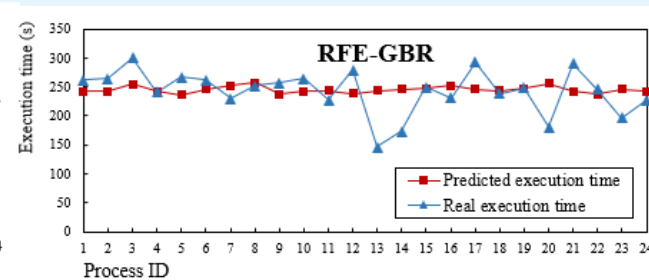
GeoIS: SPARK Processing for Raster Data



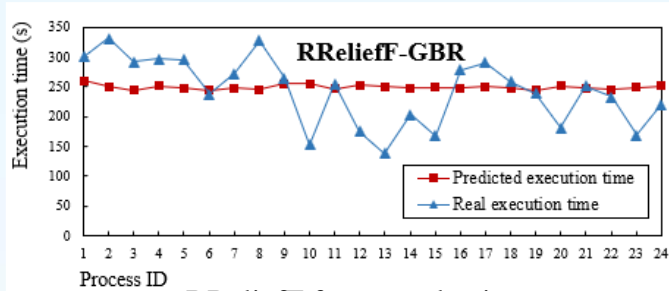
GeoIS: Machine Learning for Predicting Computational Intensity



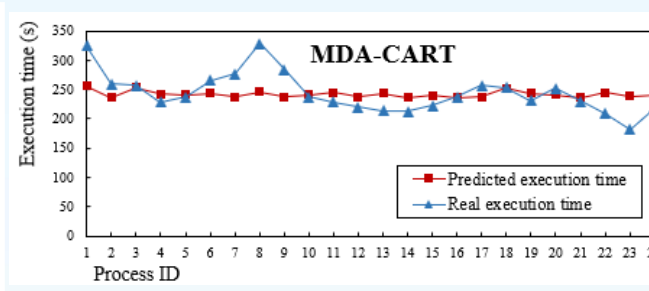
all features (AF) +
gradient boosting regressor (GBR)



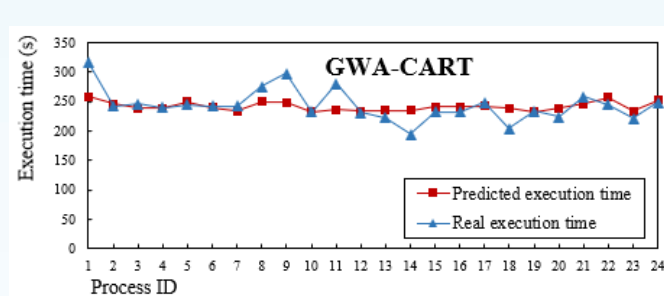
recursive feature elimination (RFE) +
gradient boosting regressor (GBR)



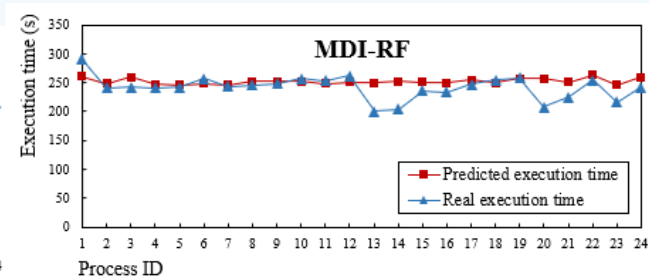
RRReliefF feature selection +
gradient boosting regressor (GBR)



mean decrease in accuracy (MDA) +
classification and regression tree (CART)



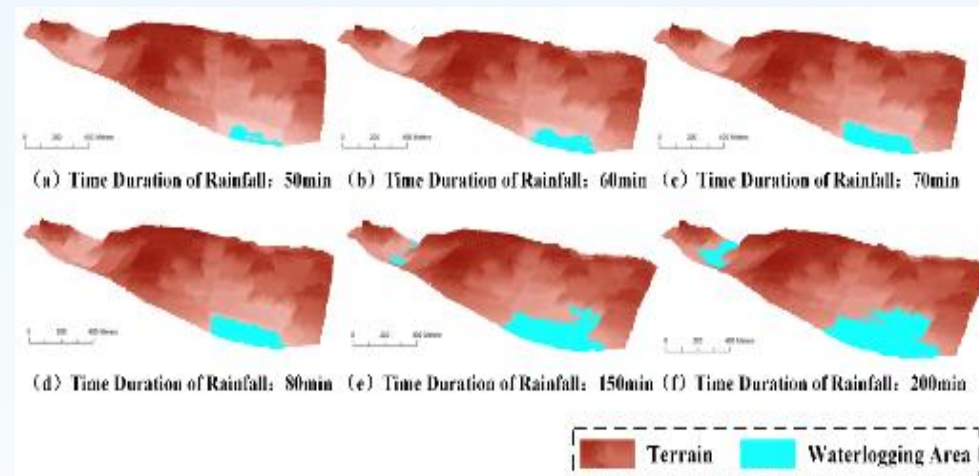
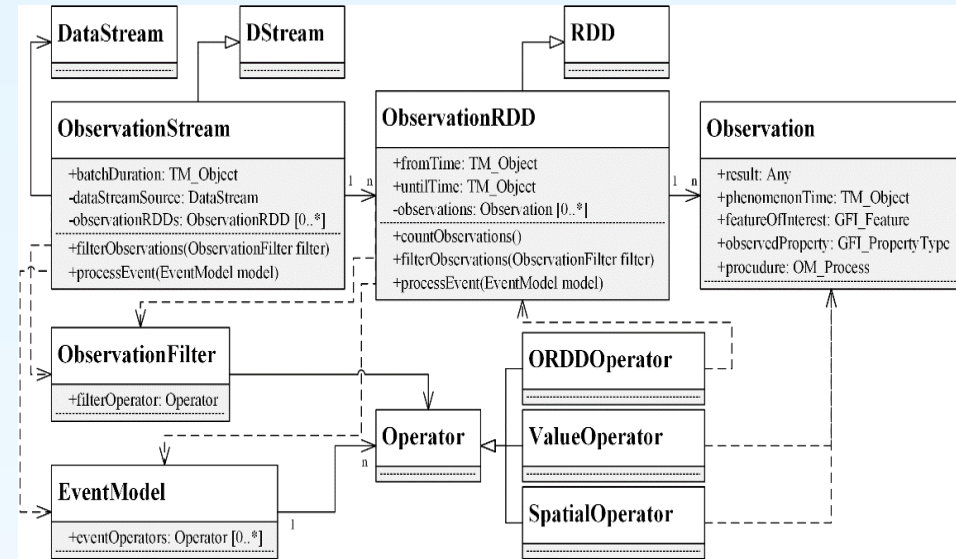
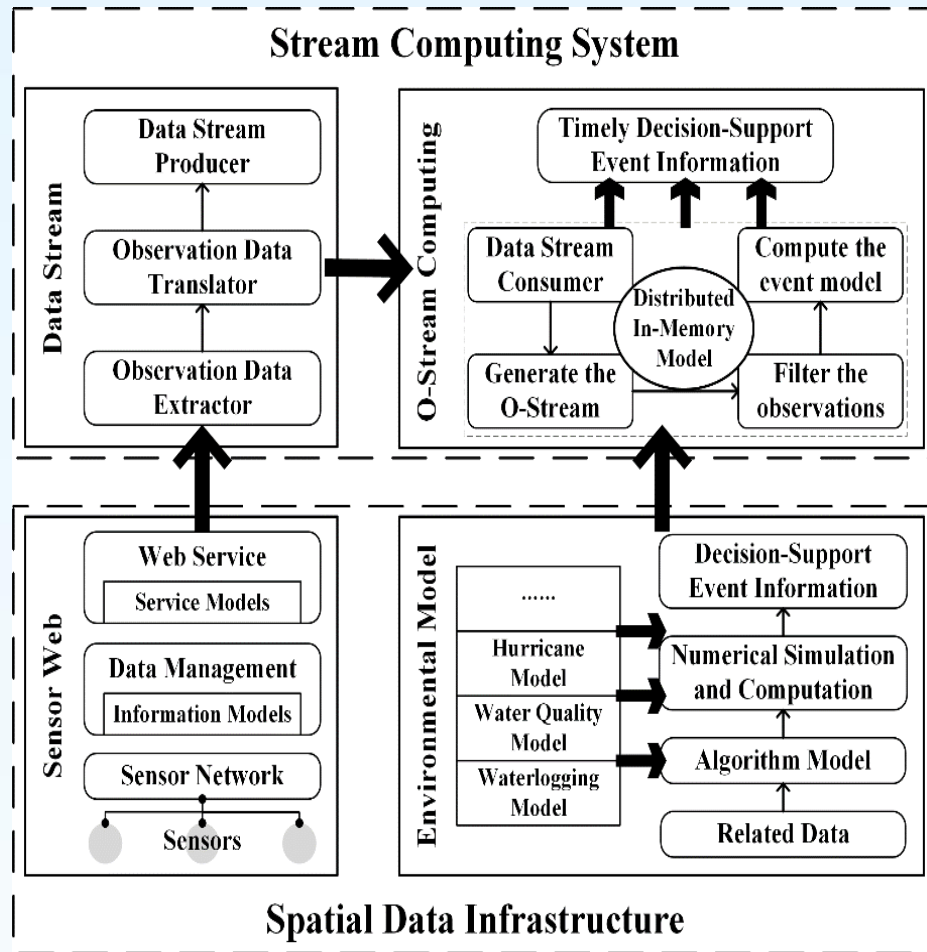
genetic-wrapper algorithm (GWA) +
classification and regression tree (CART)



mean decrease in impurity (MDI) +
random forest (RF)

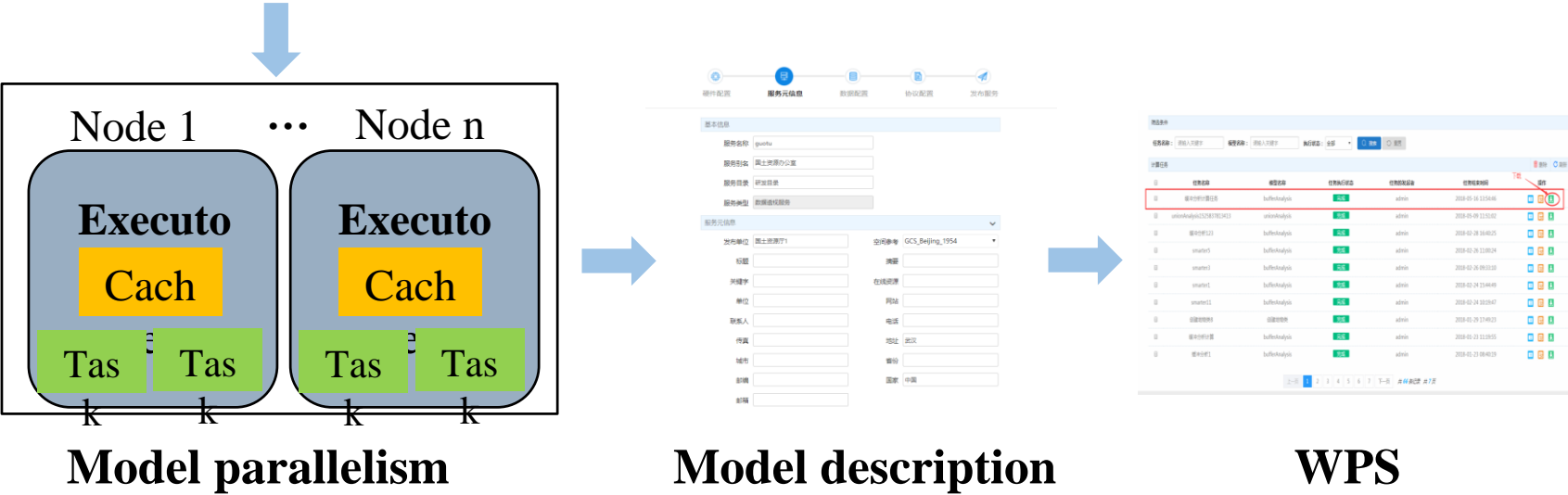
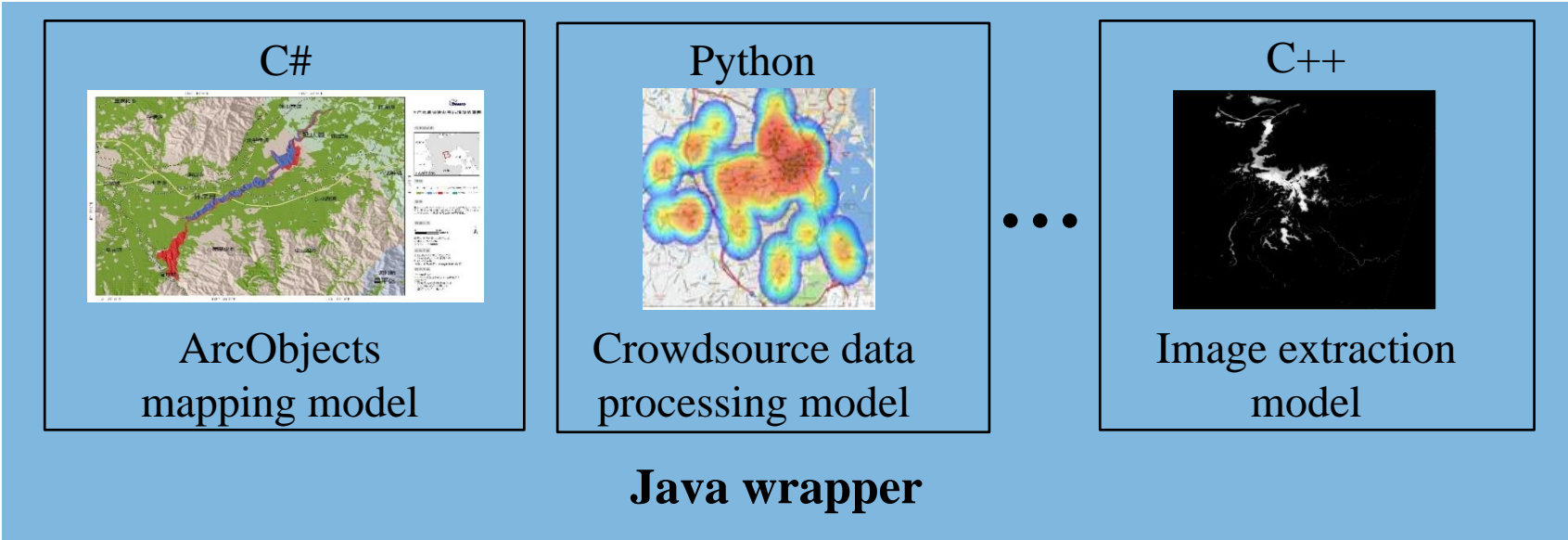
- **MDI-RF has the highest accuracy , achieve better fitting results**
- **AF-GBR, RFE-GBR, and RRReliefF has the low accuracy, and load imbalance**
- **Load balance results depends on the combination of selected features and regression models**
- **Different geoprocessing has different better feature selection algorithms and regression models**

GeoS: Stream Computing for Sensor Web observations

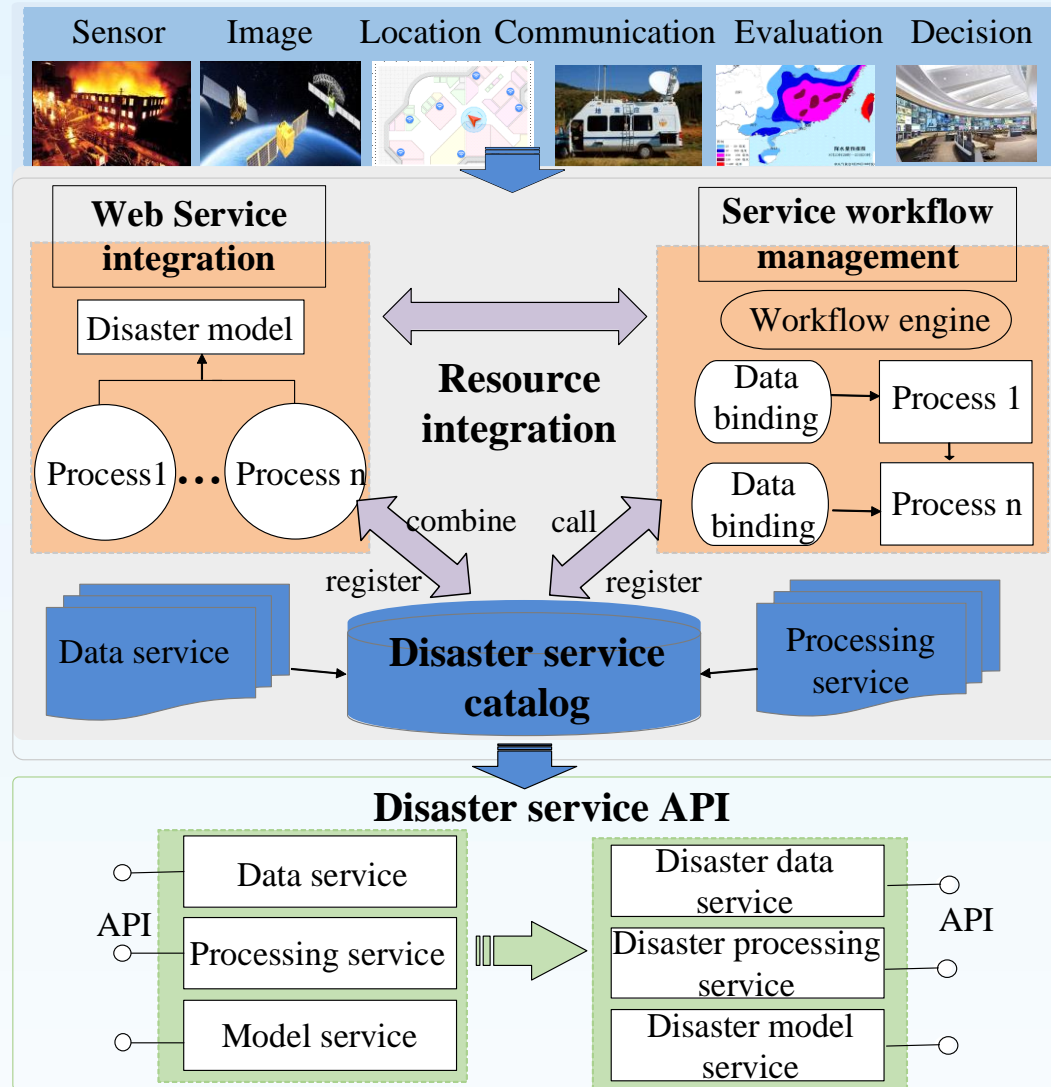


Shangguan et al., Environmental Modelling & Software (2019)

GeoS: Reusing Legacy Processing Library



GeoIS: Data and Model as Services



- **Exposing data, algorithms, and models as Services, and provide service APIs to access them**
- **Establishing a catalog service to manage services**
- **Design a workflow engine to integrate data and processing services**
- **Support customization of workflows for different scenarios**

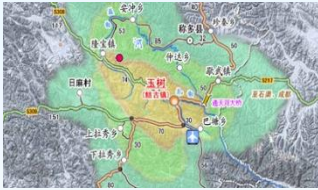
GeolS

App Store :

Managing a set of GIS projects, making engineering data sensible and accessible, reusing templates to expedite project development

Natural Disasters

Yushu Earthquake



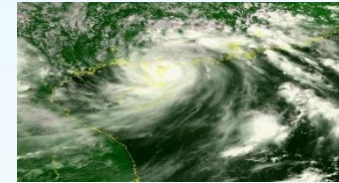
BeltRoad



Flood



Typhoon



...

Real Estates



Transportation



Smart Sales



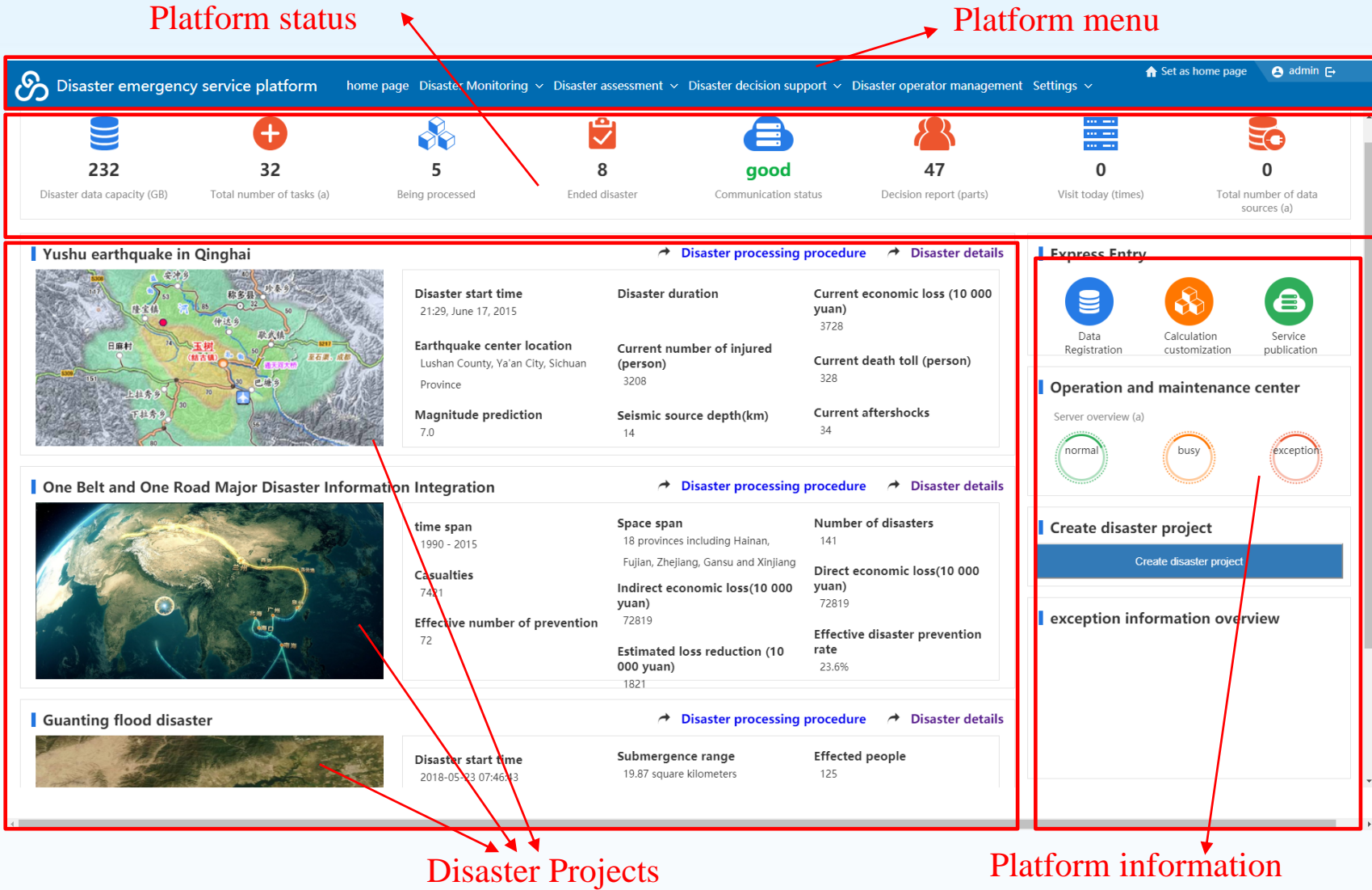
City Environment



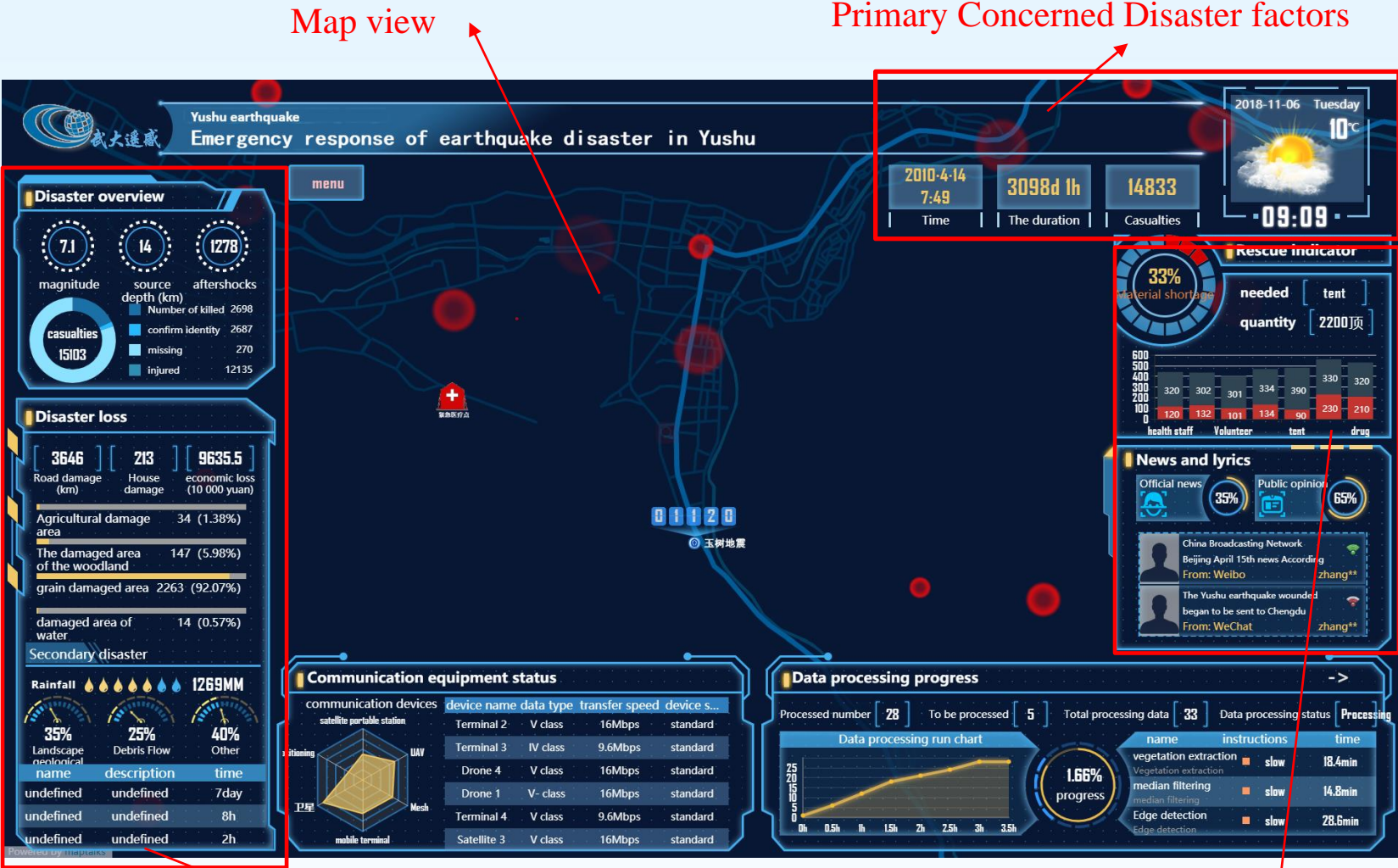
...

Smart Cities

GeoS for Natural Disaster Risk Management



Yushu Earthquake Project



Map view

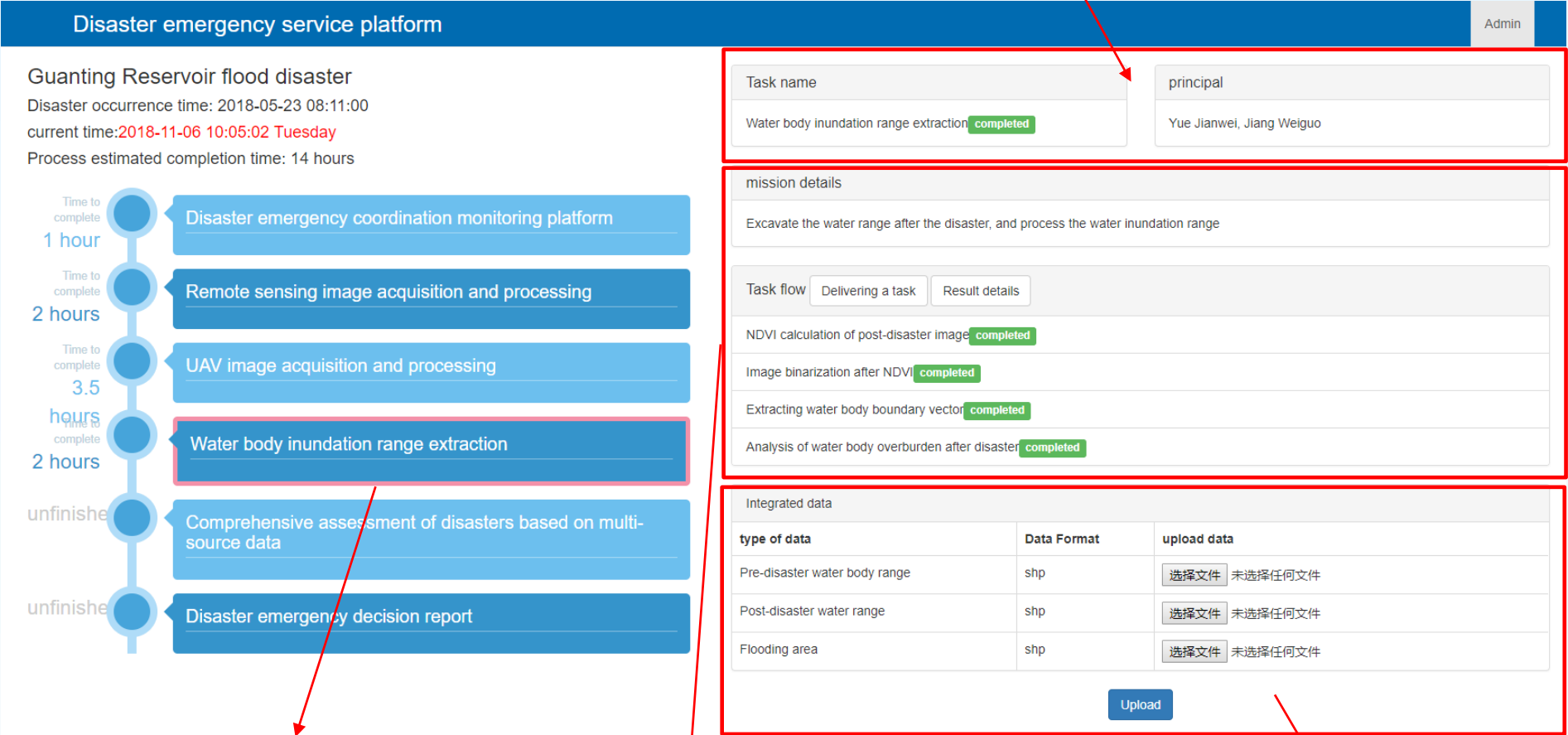
Primary Concerned Disaster factors

Disaster loss

Needs assessment/Social media

Guanting Flood Project

Disaster emergency response
Task and responsible parties

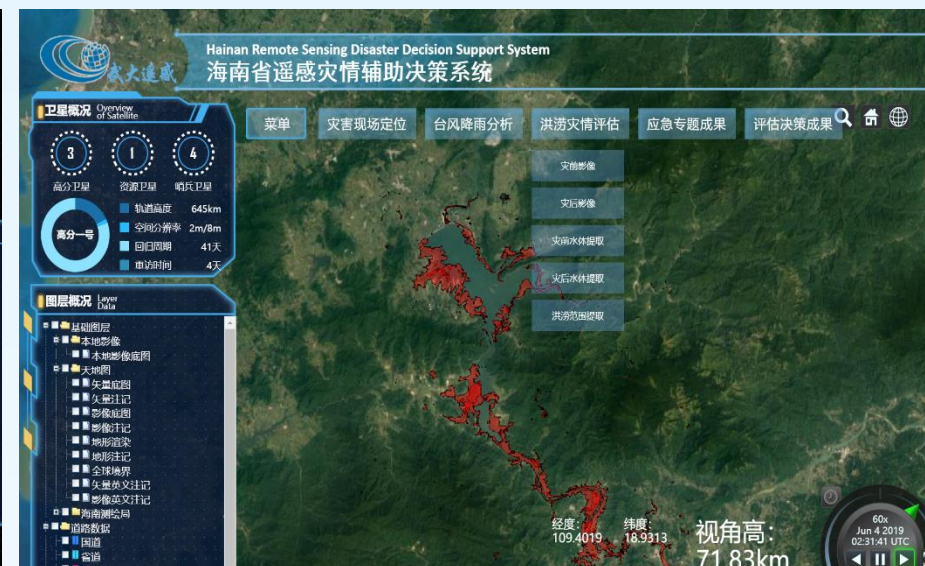
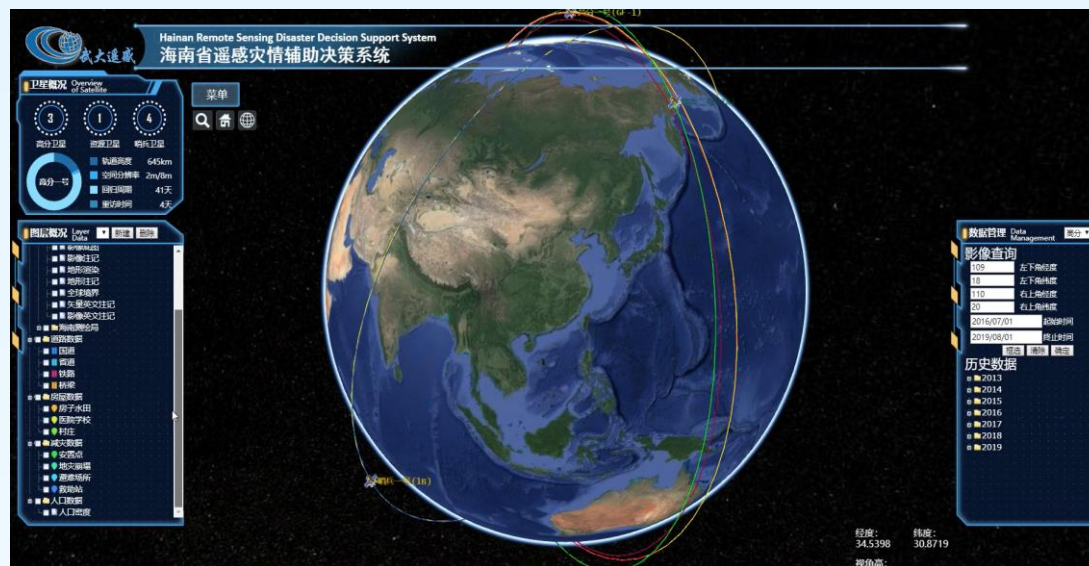


Current response step

Processing mission and status

Upload data

Provincial Remote Sensing Disaster Decision Support

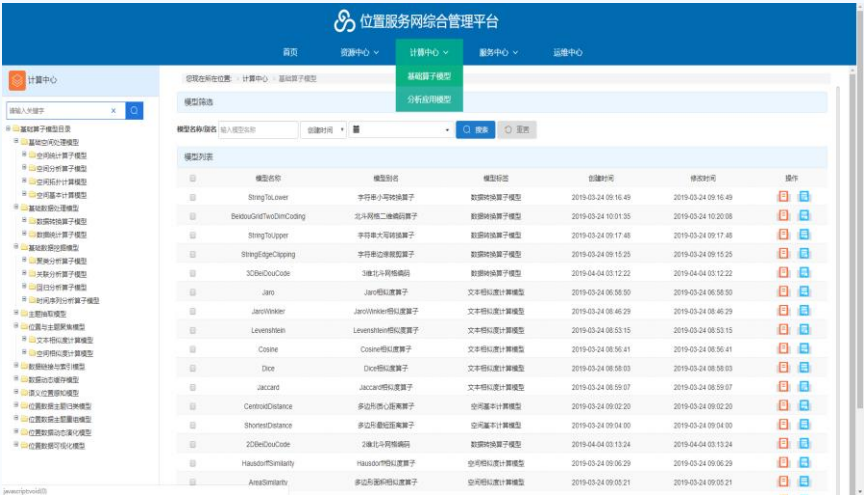
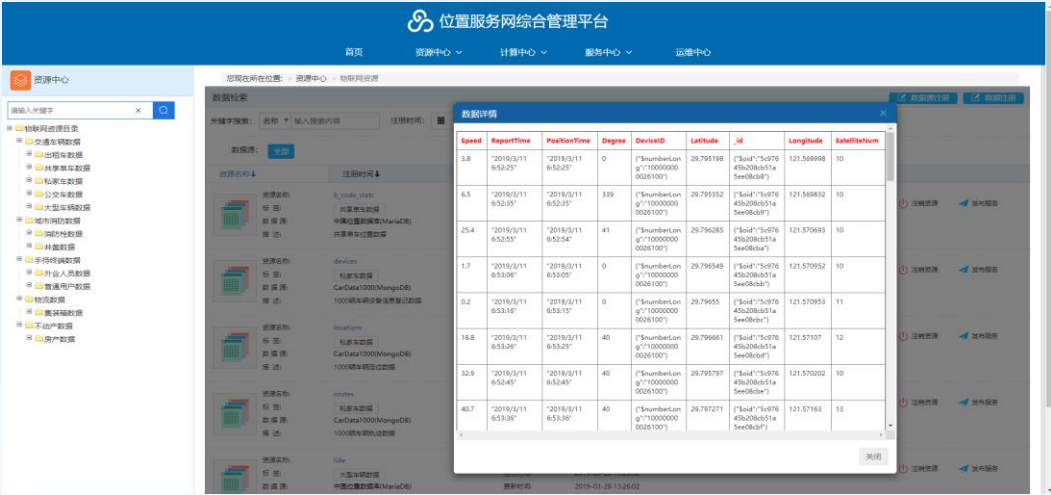


GeolS for Smart Cities



GeoS for Smart Cities

Platform Operations



GeoIS for Smart Cities

Real Estate Communities Accommodation Suitability



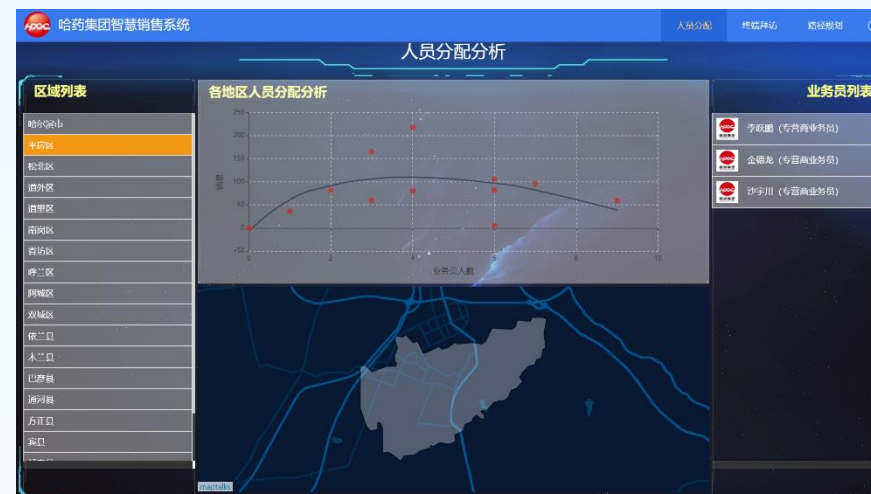
GeoS for Smart Cities

Urban Traffic Analysis



GeoIS for Smart Cities

Smart Sales (where are salesman, store, products, and profit?)



Conclusions

- **GIS engineering data is a kind of geospatial big data, which can be sensed and accessed in big data infrastructures**
- **Data, Model, App are three cores in big data enabled GIS engineering and services**
- **Better big data infrastructures, better GIS engineering and services**

Thank You!