

# Knowledge Mining for Intelligent Geospatial Data Discovery

**Wenwen Li**

GeoDa Center for Geospatial Analysis and Computation  
School of Geographical Sciences and Urban Planning  
Arizona State University  
05-07-2013

# Outline

- Background
- Intelligent geospatial data discovery
- A knowledge mining approach
- Results
- Conclusion and discussion

# From GIS to CyberGIS

- Analysis
  - Single desktop to cluster-based, cloud-based remote computing
- Data
  - Centralized database to distributed web-accessible database/catalog on the CyberSpace
- Resource discovery
  - Distributed data resources
  - Distributed analytical resources

# Current Efforts



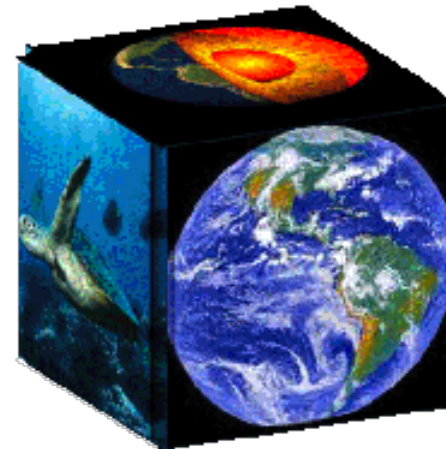
<http://inspire-geoportal.ec.europa.eu>



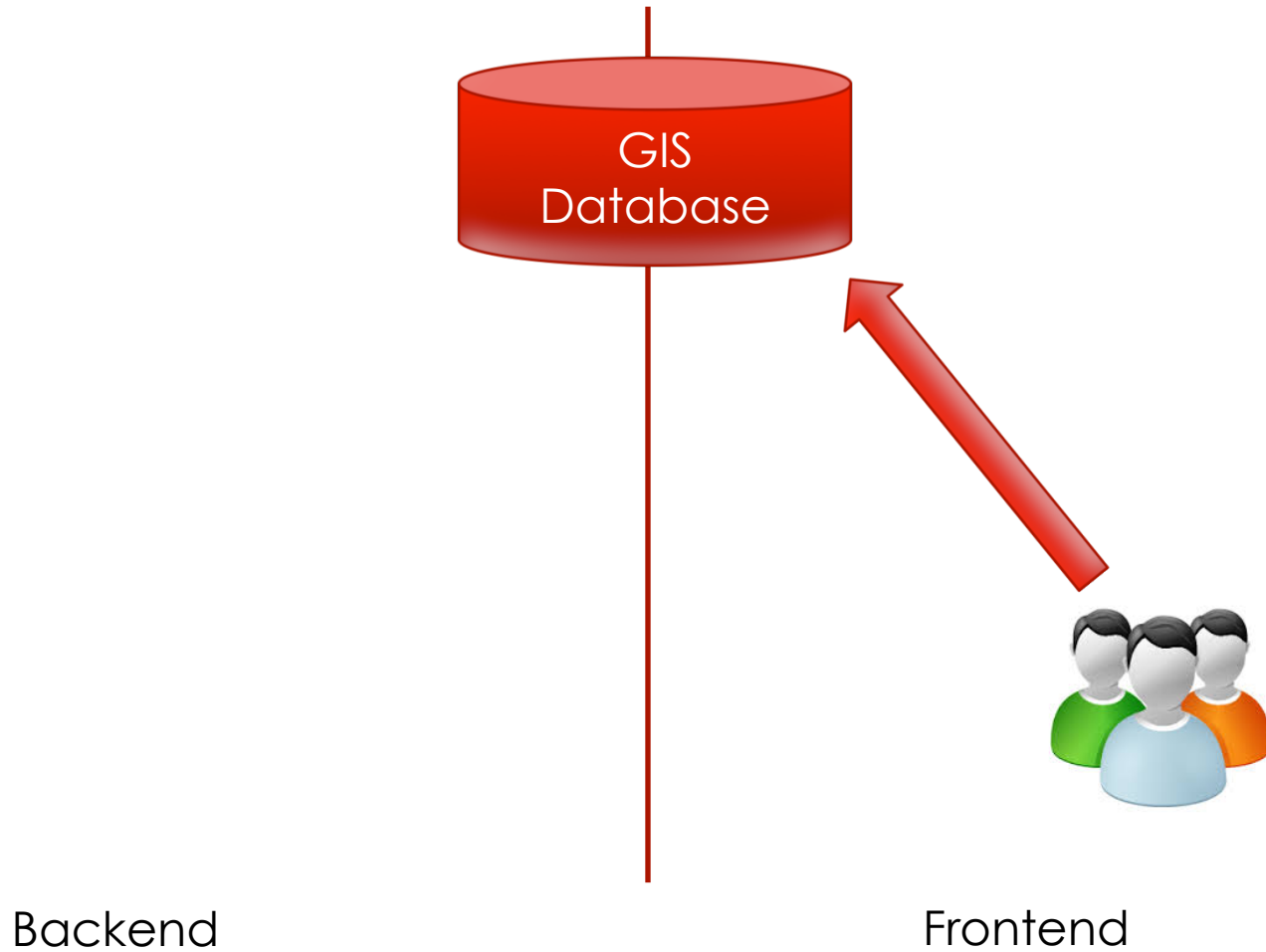
<http://www.geongrid.org>

**GeoNetwork**  
*Opensource*

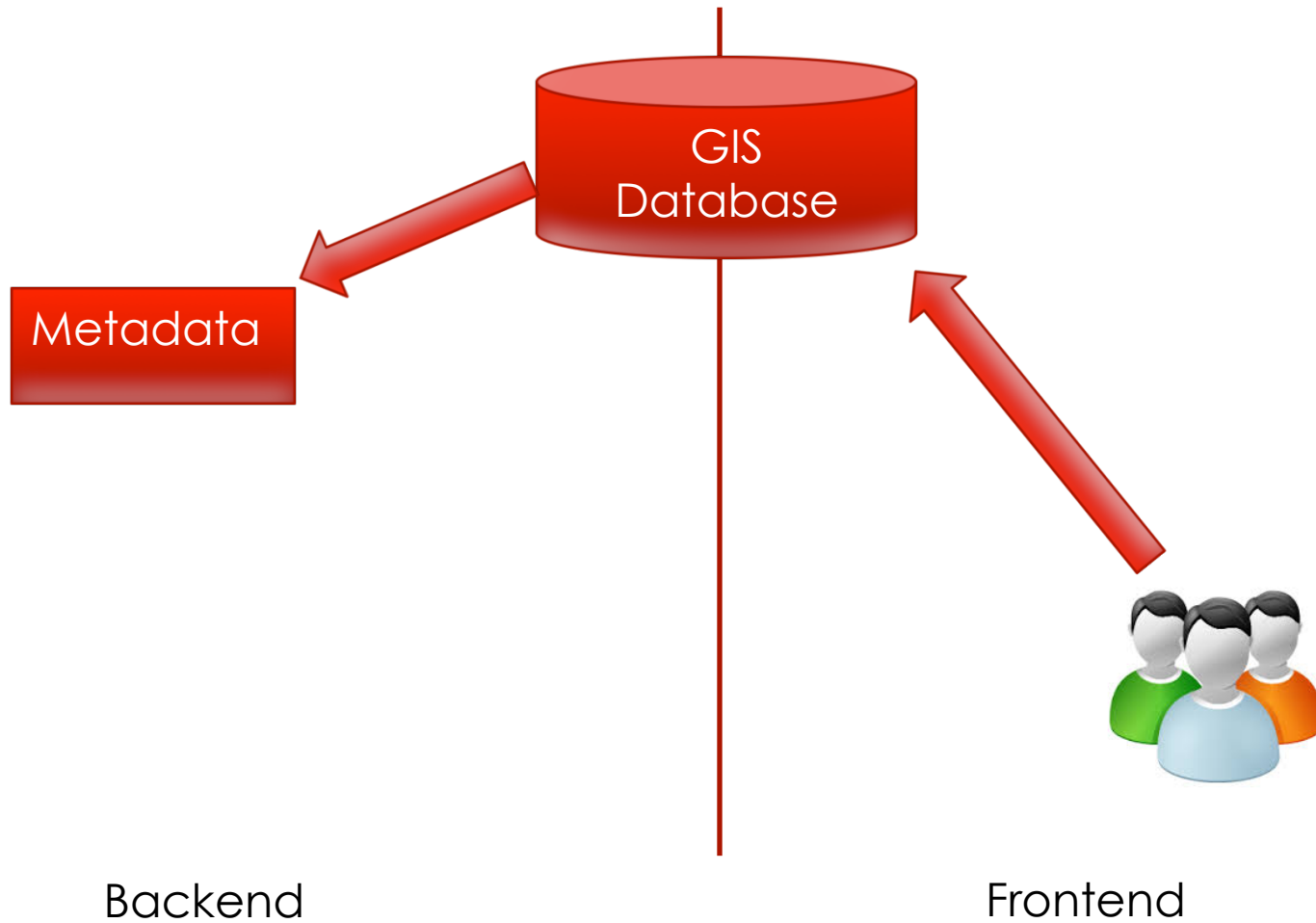
<http://geonetwork-opensource.org/>



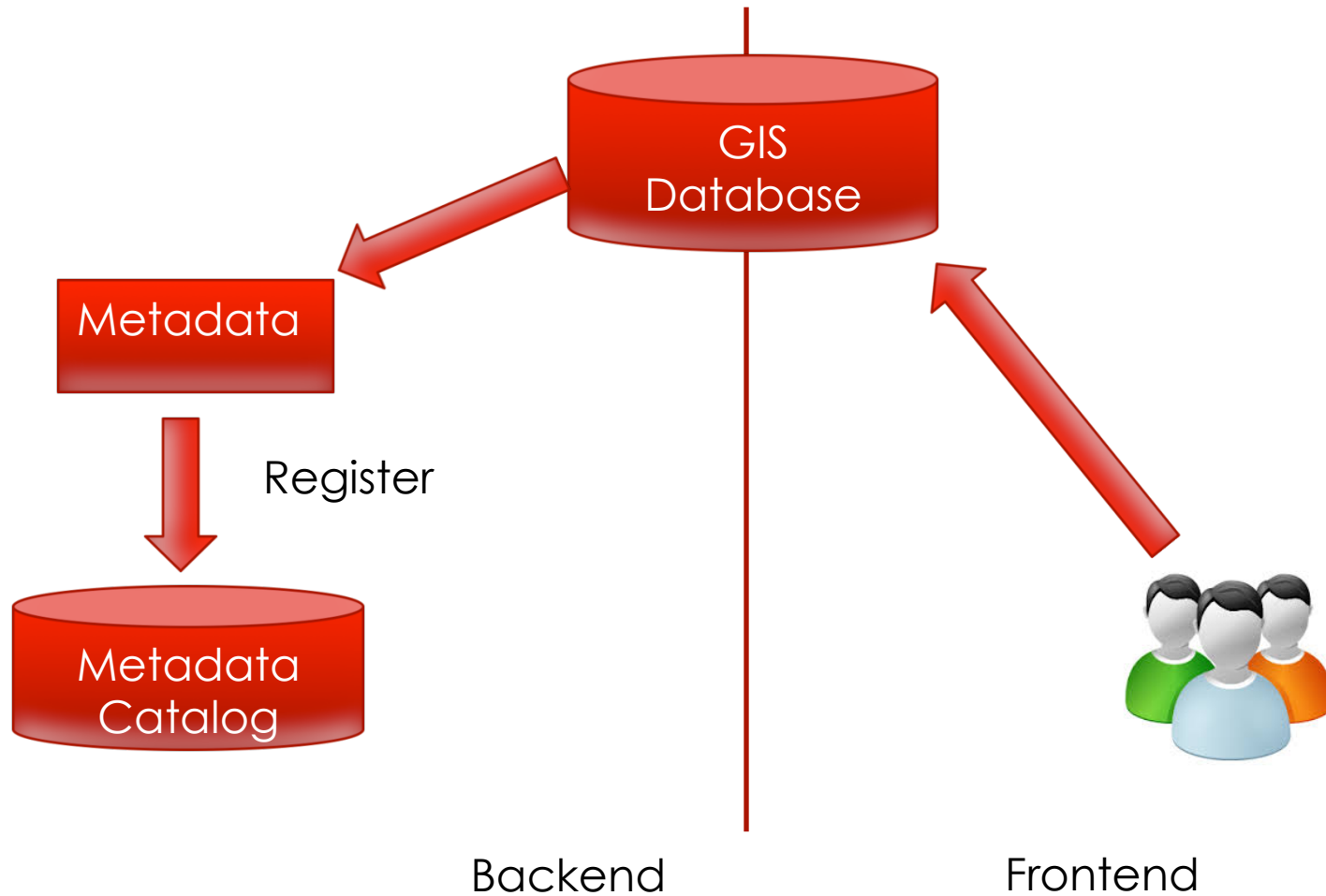
# Problem Statement



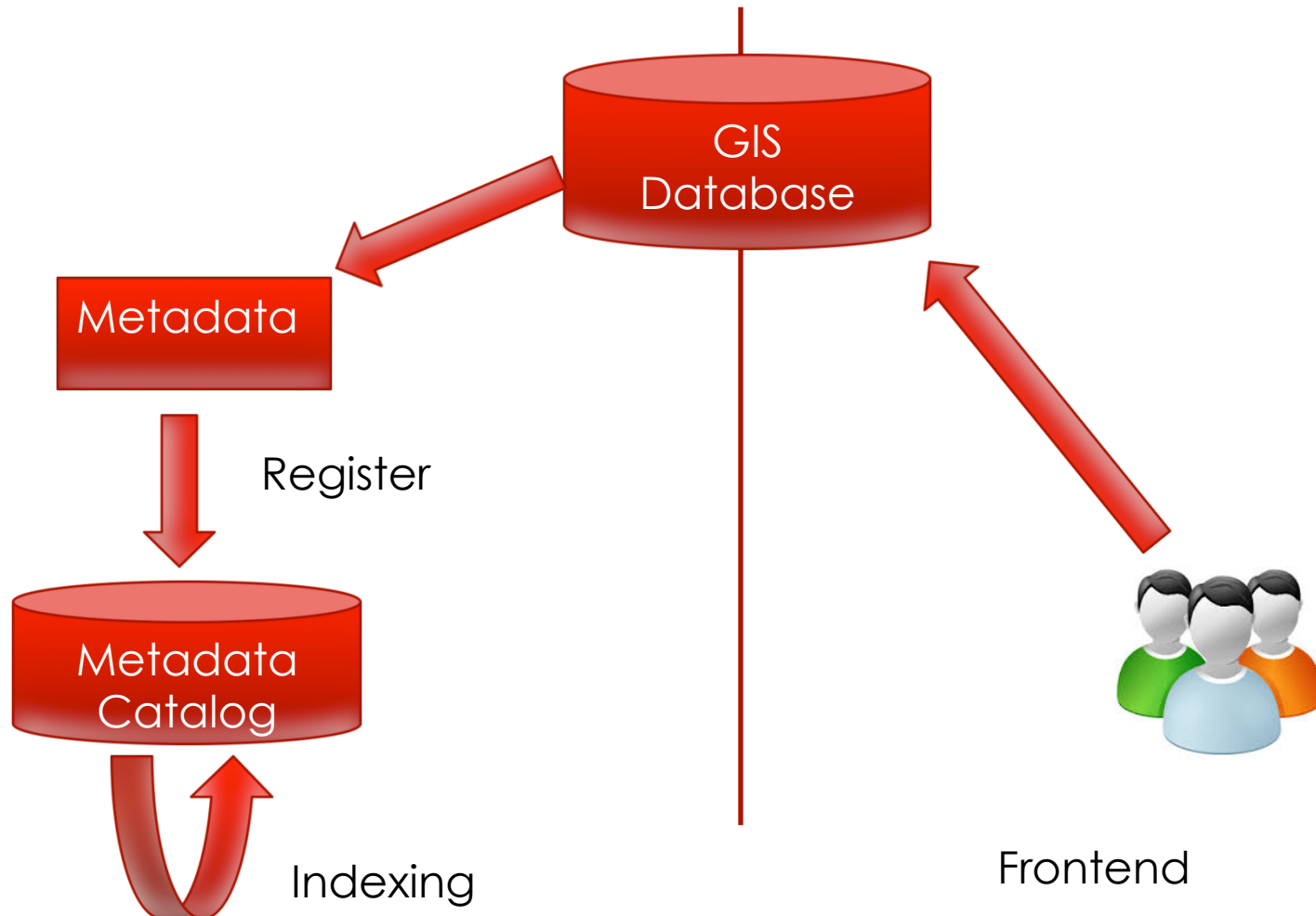
# Problem Statement



# Problem Statement

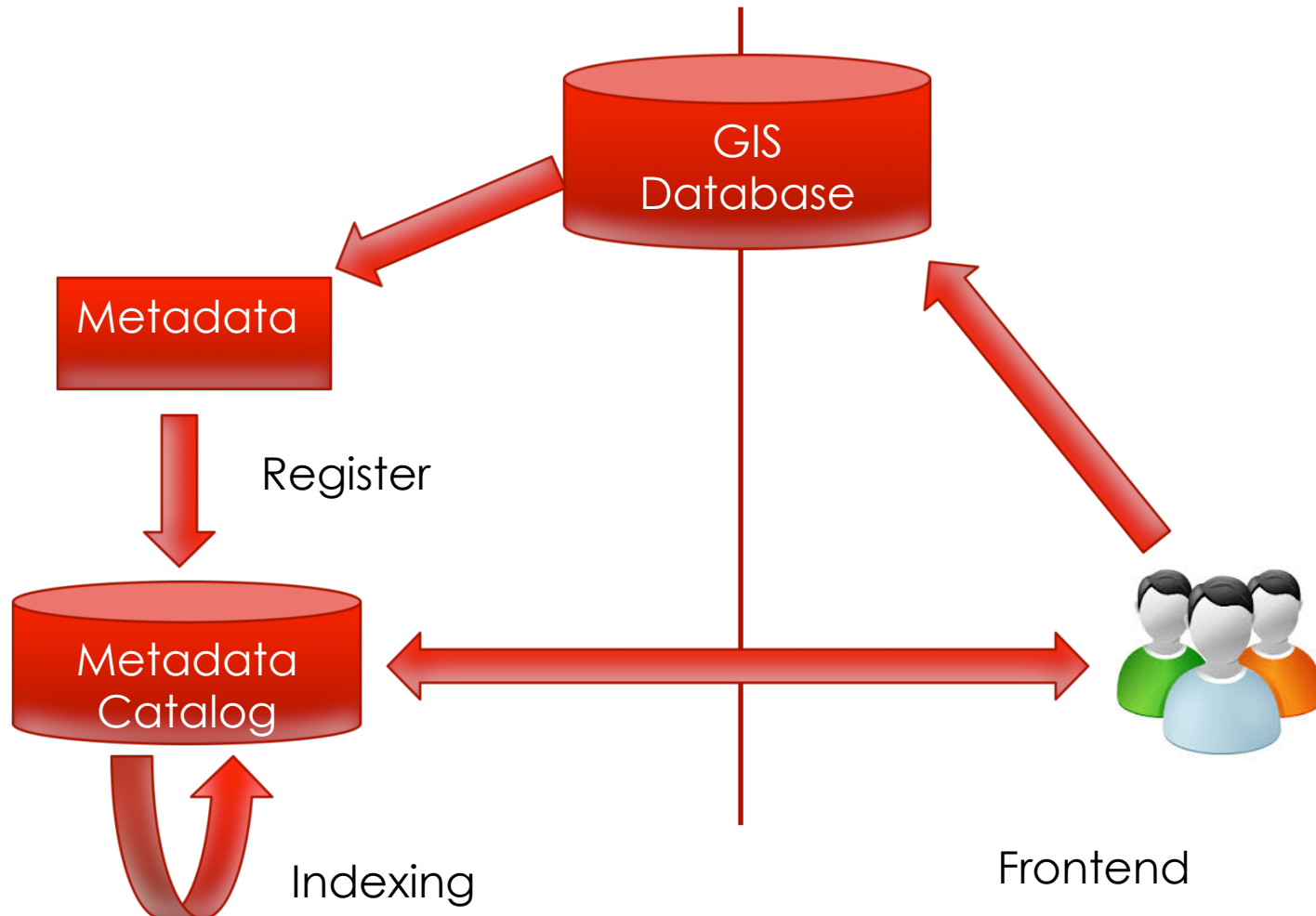


# Problem Statement

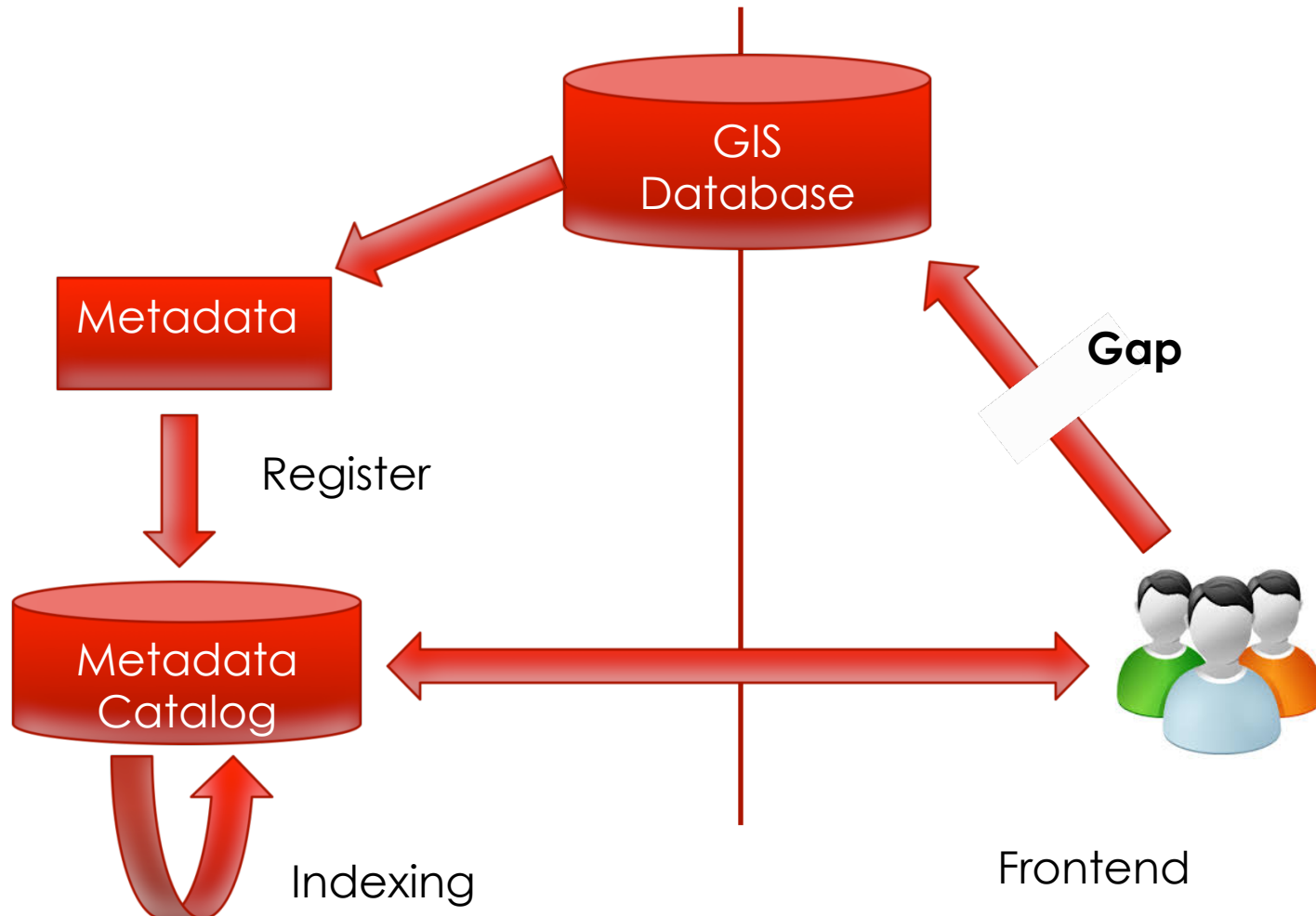




# Problem Statement



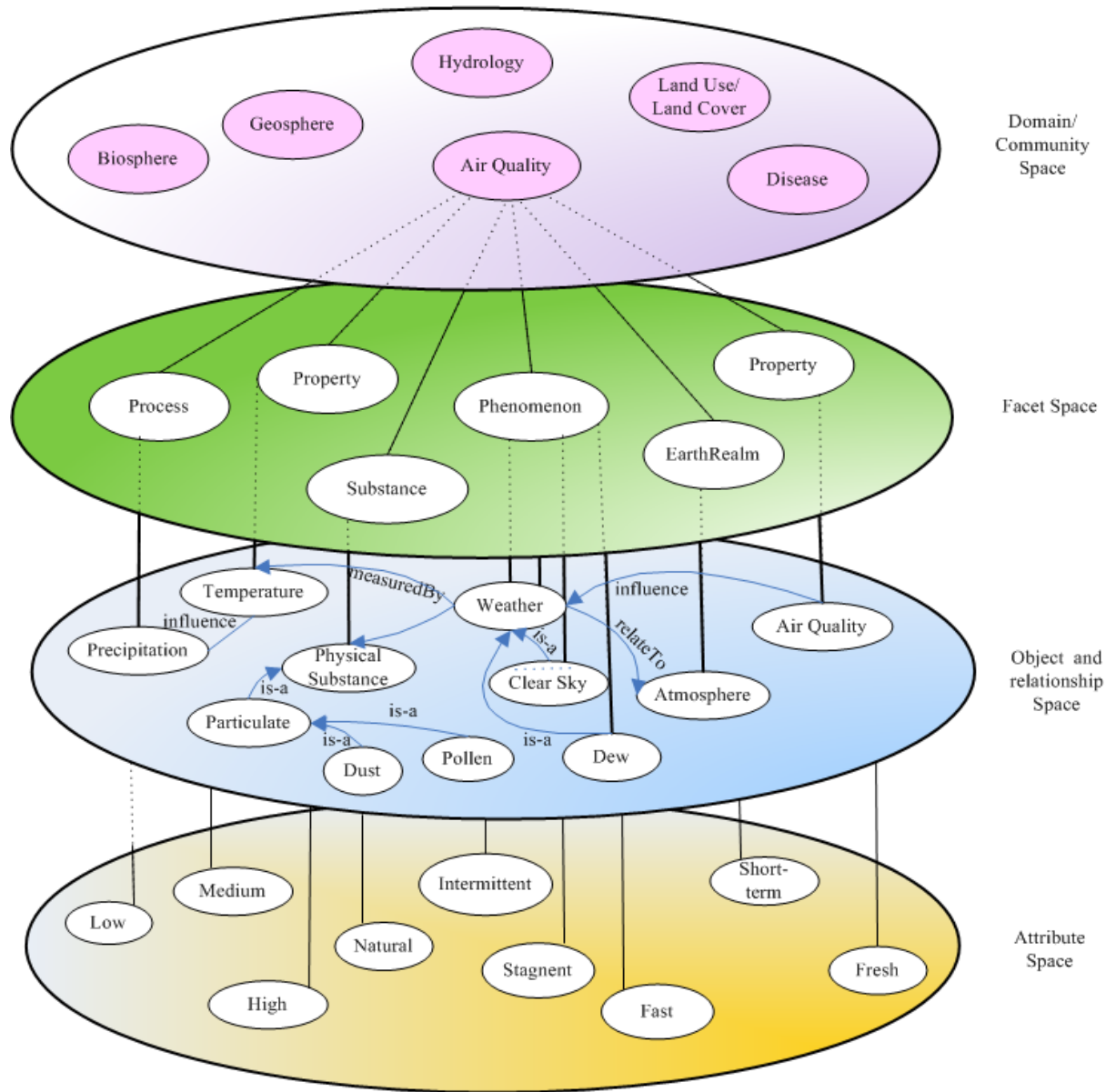
# Problem Statement



# Intelligent Geospatial Data Discovery

- Ontology-based approach
  - Formalize the definition of domain knowledge
  - Classes, individuals, properties, relationships
- Current Work:
  - GEON (Bowers et al. 2004)
  - LEAD (Droegemeler et al. 2005)
  - VSTO (Fox et al. 2008)
  - SWEET based semantic search (Li et al. 2011, Li et al. 2012a)

# A Conceptual Framework



# Semantic Search

Current Layer: phairs-devel.unm.edu

Layers: BeginTime: 2008-01-07T03:..., EndTime: 2008-01-08T03:...

Search: air quality

Web Number of results: 15 for Dust [definition].

Title: REASoN WebMap Interface  
Source: esg  
OnlineAccessible: no  
Description: WMS Server for data related to the PHAIRS REASoN project showing outputs of the DREAM eta model and elevations of atmospheric particulate matter.

Resources:  
 NOAA NCDC - 10 ✓  
 NASA GCMD - 15 ✓  
 FGDC GOS - 15 ✓  
 NASA ECHO - 0 ✓  
 NASA ESG - 2 ✓

Refine Your Search Here:  
 +Contaminant  
 +Particulate  
 +Dust

Web Number of results: 55 for air quality [definition].

Title: Ambient Air Quality Monitoring Program [gcmd\_60]  
Source: gcmd  
OnlineAccessible: yes  
Description: Summary: The EPA's ambient air quality monitoring program is carried out by State... Air Monitoring Stations (SLAMS), National Air Monitoring Stations (NAMS), and Special Purpose

Resources:  
 NOAA NCDC - 10 ✓  
 NASA GCMD - 15 ✓  
 FGDC GOS - 15 ✓  
 NASA ECHO - 0 ✓  
 NASA ESG - 0 ✓

Refine Your Search Here:  
 +Phenomena  
 +Weather  
 Front  
 Clear Sky  
 Dew  
 Hydrometeor  
 Blowing  
 Obscuration  
 Atmospheric Cir...  
 Frost

Related Terms:  
 +Phenomena  
 Precipitation  
 Cloud  
 +Property  
 Pressure  
 Temperature  
 +Earthrealm  
 Atmosphere  
 +Substance  
 PhysicalSubstance

Refine Your Search Here:  
 +Material  
 +Substance  
 +PhysicalSubstance  
 Greenhouse Gas  
 Natural Resourc...  
 Water  
 Solid Substance  
 Deposit  
 Liquid Substanc...  
 Hazardous Subst...  
 Gaseous Substan...  
 Mixed Chemical ...  
 Mixed Substance  
 Particulate

Related Terms:  
 +Phenomena  
 Atmospheric Circulat...  
 Weather  
 +Property  
 Humidity  
 ThermalConductivity  
 Velocity  
 Turbidity

# Limitation

- **Hard to model spatial relationship** (Shi, 2011)
  - Equal, within, touch, disjoin, intersect..
- **Hard to build a consensus domain ontology** (ontology mapping)

Ontology	Full Name	Creator
SWEET	Semantic Web for Earth and Environmental Terminology	NASA JPL
CUAHSI	Consortium of Universities for the Advancement of Hydrologic Science	CUAHSI
MMI	Marine Metadata Interoperability	NSF
INSPIRE	Infrastructure for Spatial Information in Europe	European Commission
GEMET	General Multilingual Environmental Thesaurus	EEA, ETC/CDS

- **Limited spatial reasoning capability**
  - Similarity reasoning
    - Rodriguez and Egenhofer, 2004; Janowicz et al. 2008; Li et al. 2012a
  - Natural language processing

# A Knowledge Mining Approach

- Goal:
  - Identify latent semantic associations rather than manually built-up
  - Search based upon meaning rather than appearance
  - Bottom-up approach: let the data speak
- Methodology:
  - Latent Semantic Analysis (LSA)

# Intro to LSA

- Mathematical approach for computer modeling and simulation of the MEANING of words and paragraphs
- Identify semantic structure of domain knowledge residing in the metadata files.
  - Concept with similar meanings
  - Similar metadata documents
- Linear Algebra:
  - SVD: Singular Value Decomposition
  - Lower-rank estimation



# Intro to LSA

**Point of contact**

Individual name: JELLE HIELKEHA  
 Organization name: UN/FAO/IS/EN/NETART >  
 METART: Environment, Sustainable Development Department, Food and Agriculture Organization, United Nations  
 Position name: DATA CENTER CONTACT  
 Role: Resource provider: Party that supplies the resource

Voice: +39 06 5705589  
 Facsimile: +39 06 57022689  
 Delivery point: Food and Agriculture Organization (FAO)  
 Delivery point: Environment and Natural Resources Service (ENRS)  
 Delivery point: Viale della Terme di Caracalla  
 City: Rome  
 Postal code: 00150  
 Country: Italy  
 Electronic mail address: [jelle.hielkeha@fao.org](mailto:jelle.hielkeha@fao.org)  
 Online resource: <http://isnetart.fao.org>

**Geographic keywords**  
 EARTH SCIENCE > ATMOSPHERE > CLOUDS > CLOUD AMOUNT/FREQUENCY >>> CLOUD COVER, EARTH SCIENCE > ATMOSPHERE > PRECIPITATION > PRECIPITATION AMOUNT >>> EARTH SCIENCE > ATMOSPHERE > PRECIPITATION > RAIN >>> RAINFALL, EARTH SCIENCE > BIOSPHERE > VEGETATION > BIOMASS >>> EARTH SCIENCE > BIOSPHERE > VEGETATION > VEGETATION COVER >>> EARTH SCIENCE > BIOSPHERE > VEGETATION > VEGETATION INDEX >>> >>>  
 AGRICULTURE AND FOOD SECURITY, CLOUD COVER, FOOD, LAND AND FRESHWATER RESOURCES, LOCUSTS, RAINFALL, VEGETATION, ARTENS > UN/FAO Africa Real Time Environmental Monitoring Using Imaging Satellites >  
 URI: CEDS, USA/NSGA, ..

**Language**  
 English  
 French  
 Spanish, Castilian

**Character set**  
 UTF8, 8-bit variable size UCS Transfer Format, based on ISO/IEC 10646

**Topic category code**  
 Climatology, meteorology, atmospheric  
 Biota

**Geographic bounding box**



Geospatial metadata in XML

	d1	d2	..	..	..	dn
k1						
k2						
.						
.						
km						

Inverted index: Term-Document Matrix

**k**: keyword  
**d**: documents

c1: The *geospatial* Web: how *geo*-browsers, social software and the Web 2.0

c2: *Geospatial semantics*: capture meanings of *spatial* information

c3: A *semantic search* engine for *spatial* Web portals

c4: Google's *spatial search* tools in the Marine *Environment* - Decision Support

m1: Darcy's *law* on *hydrology*

m2: *Hydrology* and *Water Law* - Bridging the Gap

m3: *Hydrology*: an *environmental* approach

m4: *Environmental law*: Hazardous wastes and substances

$r(\text{geo search}) = -0.3333$   
 $r(\text{geo law}) = -0.4472$

S =

2.7395	0	0	0	0	0	0	0	0
0	2.3709	0	0	0	0	0	0	0
0	0	1.6454	0	0	0	0	0	0
0	0	0	1.2380	0	0	0	0	0
0	0	0	0	1.0000	0	0	0	0
0	0	0	0	0	0.7963	0	0	0
0	0	0	0	0	0	0.0000	0	0

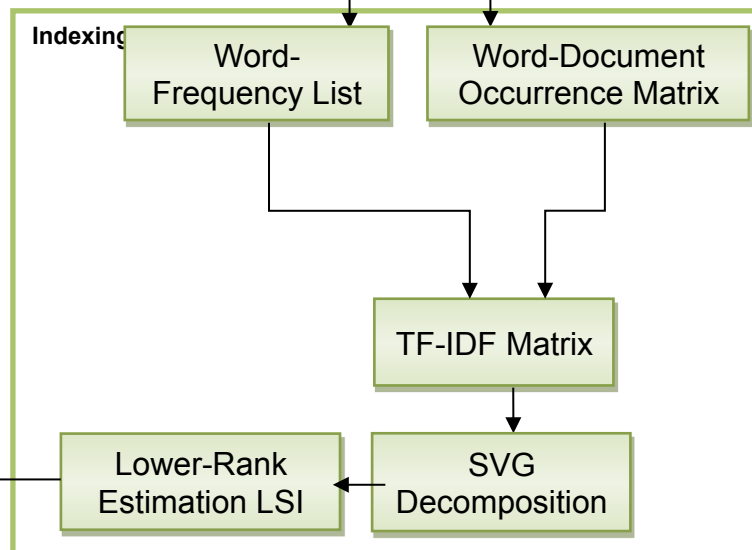
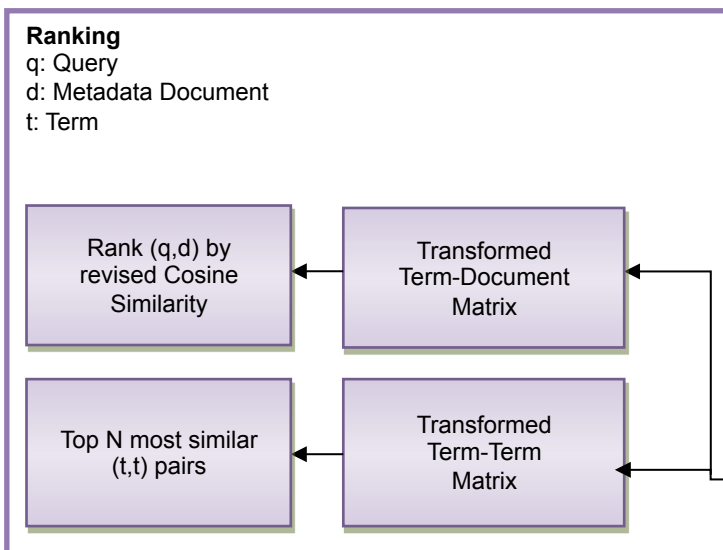
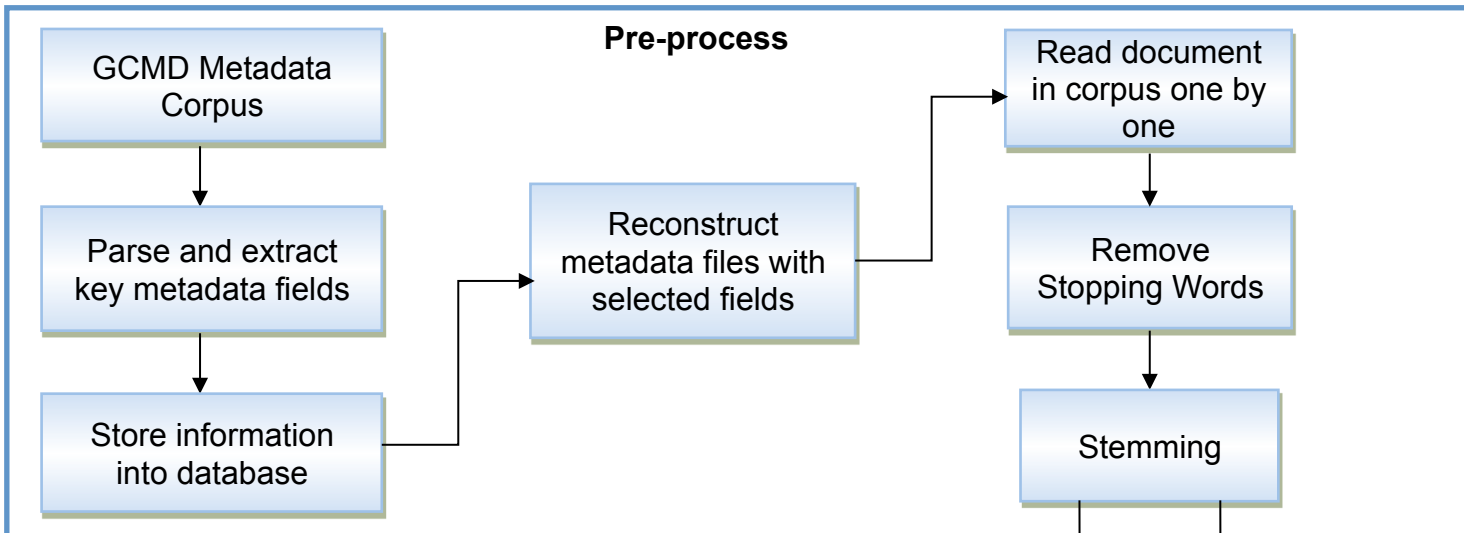
The Term-by-Document Matrix  $A$ .

Matrix $A$	c1	c2	c3	c4	m1	m2	m3	m4
<b>Geo</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
spatial	1	1	1	1	0	0	0	0
semantic	0	1	1	0	0	0	0	0
<b>search</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
Environment(al)	0	0	0	1	0	0	1	1
<b>law</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>1</b>
hydrology	0	0	0	0	1	1	1	0

Matrix $\hat{A}$	C1	C2	C3	C4	M1	M2	M3	M4
<b>Geo</b>	<b>0.3833</b>	<b>0.5307</b>	<b>0.5322</b>	<b>0.4226</b>	<b>-0.1075</b>	<b>-0.1075</b>	<b>-0.0160</b>	<b>-0.0160</b>
spatial	0.7852	1.0837	1.1007	0.9342	-0.0709	-0.0709	0.0965	0.0965
semantic	0.4459	0.6170	0.6206	0.5008	-0.1051	-0.1051	-0.0013	-0.0013
<b>search</b>	<b>0.4019</b>	<b>0.5529</b>	<b>0.5685</b>	<b>0.5116</b>	<b>0.0366</b>	<b>0.0366</b>	<b>0.1124</b>	<b>0.1124</b>
Environment(al)	0.1697	0.2210	0.2773	0.4591	0.5449	0.5449	0.5055	0.5055
<b>law</b>	<b>-0.0892</b>	<b>-0.1417</b>	<b>-0.0697</b>	<b>0.2553</b>	<b>0.7951</b>	<b>0.7951</b>	<b>0.6700</b>	<b>0.6700</b>
hydrology	-0.0892	-0.1417	-0.0697	0.2553	0.7951	0.7951	0.6700	0.6700

$r(\text{geo search}) = 0.9961$   
 $r(\text{geo law}) = -0.9655$

# System Architecture



# Geospatial Taxonomy Aided Semantic Search

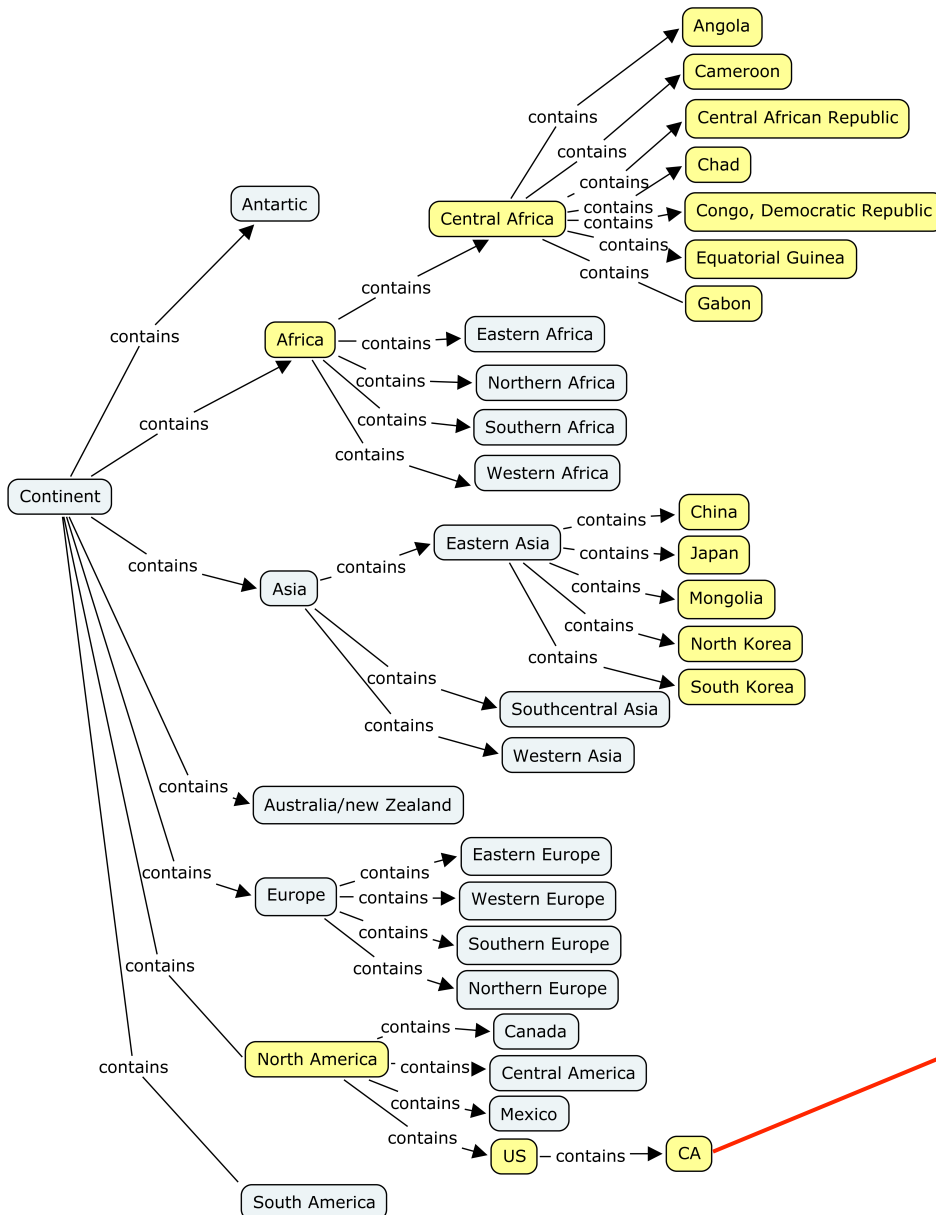
- Role:
  - Location annotation
    - Emphasize association of science keywords
    - Better handle spatial query with location with keywords
- Placename detection

```

- <gmd:descriptiveKeywords>
- <gmd:MD_Keywords>
- <gmd:keyword>
  <gco:CharacterString>CONTINENT > EUROPE > SOUTHERN EUROPE > SPAIN > GIBRALTAR > </gco:CharacterString>
</gmd:keyword>
- <gmd:keyword>
  <gco:CharacterString>CONTINENT > EUROPE > SOUTHERN EUROPE > GREECE > > </gco:CharacterString>
</gmd:keyword>
- <gmd:keyword>
  <gco:CharacterString>CONTINENT > NORTH AMERICA > GREENLAND > > > </gco:CharacterString>
</gmd:keyword>
- <gmd:keyword>
  <gco:CharacterString>OCEAN > ATLANTIC OCEAN > NORTH ATLANTIC OCEAN > CARIBBEAN SEA > GRENADA > </gco:CharacterString>
</gmd:keyword>
- <gmd:keyword>
  <gco:CharacterString>OCEAN > ATLANTIC OCEAN > NORTH ATLANTIC OCEAN > CARIBBEAN SEA > GUADELOUPE > </gco:CharacterString>
</gmd:keyword>
- <gmd:keyword>
  <gco:CharacterString>OCEAN > PACIFIC OCEAN > CENTRAL PACIFIC OCEAN > GUAM > > > </gco:CharacterString>
</gmd:keyword>
- <gmd:keyword>
  <gco:CharacterString>CONTINENT > NORTH AMERICA > CENTRAL AMERICA > GUATEMALA > > </gco:CharacterString>
</gmd:keyword>
- <gmd:keyword>
  <gco:CharacterString>CONTINENT > AFRICA > WESTERN AFRICA > GUINEA > > > </gco:CharacterString>
</gmd:keyword>
- <gmd:keyword>
  <gco:CharacterString>OCEAN > ATLANTIC OCEAN > NORTH ATLANTIC OCEAN > CARIBBEAN SEA > HAITI > </gco:CharacterString>
</gmd:keyword>
</gmd:MD_Keywords>
</gmd:descriptiveKeywords>

```

# GCMD Location Taxonomy



```
- <boundingBox>  
- <southWest>  
  <latitude>32.534290</latitude>  
  <longitude>-124.409622</longitude>  
</southWest>  
- <northEast>  
  <latitude>42.009460</latitude>  
  <longitude>-114.130836</longitude>  
</northEast>  
</boundingBox>
```

# Experimental Settings

- Benchmark

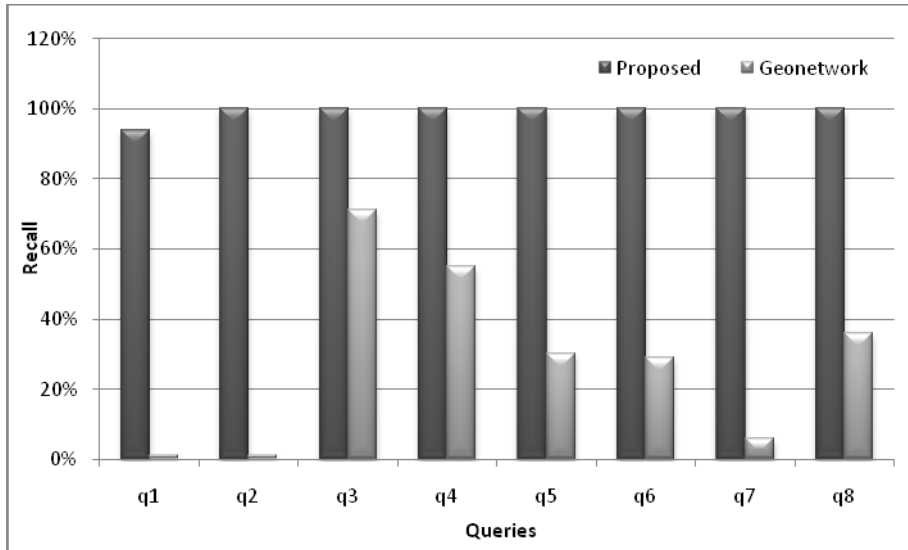
$$\text{Recall} = \frac{|\{\text{all relevant records}\} \cap \{\text{all retrieved records}\}|}{|\{\text{all relevant records}\}|}$$

$$\text{Precision} = \frac{|\{\text{all relevant records}\} \cap \{\text{all retrieved records}\}|}{|\{\text{all retrieved records}\}|}$$

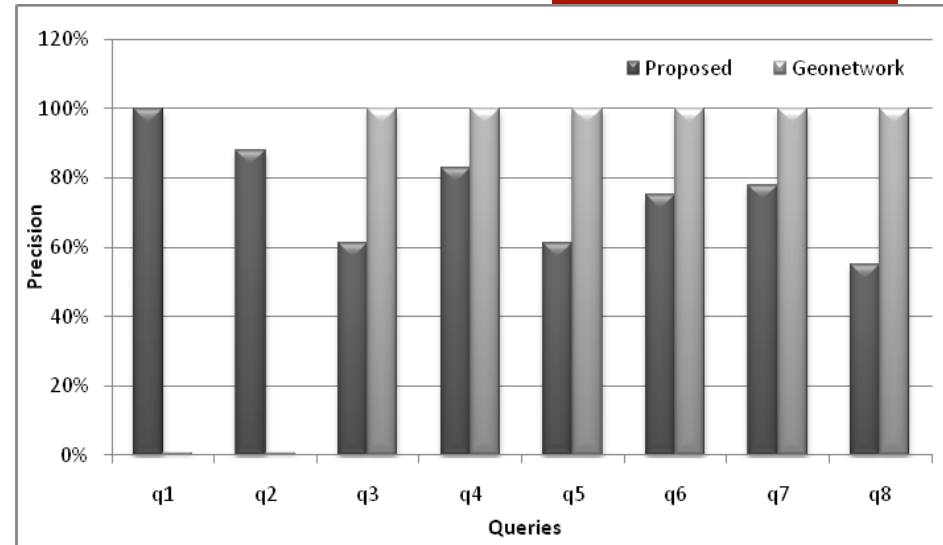
Query Type	Query	Keyword
1	Q 1.1	natural disaster death
	Q 1.2	disaster population impact
	Q 1.3	natural disaster damage
	Q 1.4	wildlife distributions by species
	Q 1.5	global climate change pollution
	Q 1.6	China agriculture food sustainability
	Q 1.7	census housing condition
	Q 1.8	Africa poverty statistics
2	Q 2.1	Colorado population
	Q 2.2	California population dynamics in the United States
	Q 2.3	wild life habitat of Costa Rica
	Q 2.4	China County level population data
	Q 2.5	Puerto Rico census data

# Results

### Recall



### Precision



- Q1.1 natural disaster death
  - Global earthquake/flood/volcano/drought/landslide/cyclone mortality
- Q1.5 global climate change pollution
  - SIR: 33 (20 related); Geonetwork: 6
  - Exception: "Global Multi-hazard total economic loss", "environmental protection" -> pollution  
cause: "hazard"

**WHAT?**  
California population dynamics

**WHERE?**



- Any -

Search

Reset Advanced Options

- Applications
- Audio/Video
- Case studies, best practices
- Conference proceedings
- Datasets
- Directories
- Interactive resources
- Maps & graphics
- Other information resources
- Photo

Show map

**FIND INTERACTIVE MAPS, GIS DATASETS, SATELLITE IMAGERY AND RELATED APPLICATIONS**

Aggregated results matching search criteria : 1-6/6 (Page1/1), 0 selected

Select : all, none

# actions on selection

Sort by Relevance

**CHINA DIMENSIONS DATA COLLECTION: CHINA COUNTY-LEVEL DATA ON POPULATION (CENSUS) AND AGRICULTURE, KEYED TO 1:1M GIS MAP**

Abstract  
Keywords

China County-Level Data on Population (Census) and Agriculture, Keyed To 1:1M GIS Map consists of census, agricultural economic, and boundary data for the administr...

EARTH SCIENCE > AGRICULTURE > AGRICULTURAL CHEMICALS > FERTILIZERS > > >, EARTH SCIENCE > AGRICULTURE > AGRICULTURAL PLANT SCIENCE > CROP/PLANT YIELDS > > >, EARTH SCIENCE > AGRICULTURE > AGRICULTURAL PLANT SCIENCE > CROPPING SYSTEMS > > >, EARTH SCIENCE > AGRICULTURE > AGRICULTURAL PLANT SCIENCE > IRRIGATION > > >, EARTH SCIENCE > AGRICULTURE > ANIMAL COMMODITIES > LIVESTOCK PRODUCTS > > >, EARTH SCIENCE > AGRICULTURE > ANIMAL SCIENCE > ANIMAL MANAGEMENT SYSTEMS > > >, EARTH SCIENCE > AGRICULTURE > ANIMAL SCIENCE > ANIMAL YIELDS > > >, EARTH SCIENCE > HUMAN DIMENSIONS > BOUNDARIES > ADMINISTRATIVE DIVISIONS > > >, EARTH SCIENCE > HUMAN DIMENSIONS > BOUNDARIES > POLITICAL DIVISIONS > > >, EARTH SCIENCE > HUMAN DIMENSIONS > POPULATION > POPULATION DISTRIBUTION > > >, EARTH SCIENCE > HUMAN DIMENSIONS > POPULATION > POPULATION SIZE > > >, CIESIN > Agriculture and Food Security > Agricultural Production > Agricultural Outputs, CIESIN > Agriculture and Food Security > Agricultural Production > Agricultural Resources, CIESIN > Agriculture and Food Security > Agricultural Production > Animal Yields, CIESIN > Agriculture and Food Security > Agricultural Production > Crop Yields, CIESIN > Agriculture and Food Security > Agricultural Production > Irrigated Agriculture, CIESIN > Agriculture and Food Security > Agricultural Production > Mechanized Agriculture, CIESIN > Economic Activity > Labor > Labor Force, CIESIN > Human Attitudes, Preferences, and Behavior > Literacy > Literacy Rates, CIESIN > Industry and Energy > Industrial Production > Industrial Activities, CIESIN > Population Dynamics > Demographic Characteristics > Age, CIESIN > Population Dynamics > Demographic Characteristics > Education Level, CIESIN > Population Dynamics > Demographic Characteristics > Ethnicity, CIESIN > Population Dynamics > Demographic Characteristics > Gender, CIESIN > Population Dynamics > Demographic Characteristics > Marital Status, CIESIN > Population Dynamics > Demographic Characteristics > Occupation, CIESIN > Population Dynamics > Demographic Characteristics > Place of Residence, CIESIN > Population Dynamics > Demographic Characteristics > Population Distribution, CIESIN > Population Dynamics > Demographic Characteristics > Population Size, CIESIN > Population Dynamics > Human Settlements > Industrial Areas, CIESIN > Population Dynamics > Human Settlements > Rural Areas, CIESIN > Population Dynamics > Human Settlements > Urbanization, CIESIN > Population Dynamics > Vital Statistics > Mortality, Age-Sex Distribution, Animal Husbandry, Birth Count, Cultivated Areas, Death Count, Fishing, Forestry, GIS, Immigration, Rural Industry, Maps, Rural Population, Urban-Rural Residence, ESIP, ESIP > Earth Science Information Partners Program, CD > China Dimensions, EOSDIS > Earth Observing System Data Information System, USA/CIESIN, ECHO

Metadata

Delete Other actions

**CHINA DIMENSIONS DATA COLLECTION: AGRICULTURAL STATISTICS OF THE PEOPLE'S REPUBLIC OF CHINA: 1949-1990**

Abstract  
Keywords

Agricultural Statistics of the People's Republic of China, 1949-1990 is an historical collection of agricultural statistical data compiled by China's State Statist...

EARTH SCIENCE > AGRICULTURE > ANIMAL COMMODITIES > DAIRY PRODUCTS > > >, EARTH SCIENCE > AGRICULTURE > ANIMAL COMMODITIES > LIVESTOCK PRODUCTS > > >, EARTH SCIENCE > AGRICULTURE > ANIMAL COMMODITIES > POULTRY PRODUCTS > > >, EARTH SCIENCE > AGRICULTURE > PLANT COMMODITIES > FRUIT PRODUCTS > > >, EARTH SCIENCE > AGRICULTURE > PLANT COMMODITIES > HORTICULTURAL PRODUCTS > > >, EARTH SCIENCE > AGRICULTURE > PLANT COMMODITIES > VEGETABLE PRODUCTS > > >, EARTH SCIENCE > BIOSPHERE > ECOLOGICAL DYNAMICS > ECOSYSTEM FUNCTIONS > CONSUMPTION RATES > > >, EARTH SCIENCE > HUMAN DIMENSIONS > ATTITUDES,PREFERENCES,BEHAVIOR > CONSUMER BEHAVIOR > > >, EARTH SCIENCE > HUMAN DIMENSIONS > ATTITUDES,PREFERENCES,BEHAVIOR > SOCIAL BEHAVIOR > > >, EARTH SCIENCE > HUMAN DIMENSIONS > ECONOMIC RESOURCES > AGRICULTURAL ECONOMICS > > >, EARTH SCIENCE > HUMAN DIMENSIONS > LAND USE/LAND COVER > LAND MANAGEMENT > > >, EARTH SCIENCE > AGRICULTURE > PLANT COMMODITIES > FIELD CROP PRODUCTS > > >, CIESIN > Agriculture and Food Security > Agricultural Production > Agricultural Chemicals, CIESIN > Agriculture and Food Security > Agricultural Production > Agricultural Commodities, CIESIN > Agriculture and Food Security > Agricultural Production > Agricultural Inputs, CIESIN > Agriculture and Food Security > Agricultural Production > Agricultural Outputs, CIESIN > Agriculture and Food Security > Agricultural Production > Agricultural Resources, CIESIN > Economic Activity > Financial Policy > Credit, CIESIN > Economic Activity > Financial Policy > Expenditures, CIESIN > Economic Activity > Financial Policy > Loans, CIESIN > Economic Activity > Income > National Income, CIESIN > Economic Activity > Labor > Labor Force, CIESIN > Economic Activity > Markets > Price Index, CIESIN > Economic Activity > Trade > Goods, CIESIN > Economic Activity > Trade > Services, CIESIN > Environmental Protection > Land Use > Agricultural Use, CIESIN > Human Attitudes, Preferences, and Behavior > Consumer Behavior > Consumption, CIESIN > Industry and Energy > Energy Production > Energy Outputs, CIESIN > Industry and Energy > Industrial Production > Industrial Outputs, Agricultural Acreage, Agricultural Investment, Costs of Production, Deposits, Economic Indicators, Livestock Inventory, Price Index, Social Indicators, ESIP, ESIP > Earth Science Information Partners Program, CD > China Dimensions, EOSDIS > Earth Observing System Data Information System, USA/CIESIN, ECHO

Metadata

Delete Other actions

**CHINA DIMENSIONS DATA COLLECTION: GUOBBIAO (GB) CODES FOR THE ADMINISTRATIVE DIVISIONS OF THE PEOPLES REPUBLIC OF CHINA**

Abstract  
Keywords

GuoBiao (GB) Codes for the Administrative Divisions of the People's Republic of China consists of geographic codes for the administrative divisions of China. The d...

EARTH SCIENCE > HUMAN DIMENSIONS > BOUNDARIES > ADMINISTRATIVE DIVISIONS > > >, EARTH SCIENCE > HUMAN DIMENSIONS > BOUNDARIES > POLITICAL DIVISIONS > > >, CIESIN

Metadata

Delete Other actions



## Q 2.2 California population dynamics in the United States

LSI Model

**Spatial-Awareness in Latent Semantic Analysis**

Take a look at **Q uery Vector Coordinates** tab for query results

Enter the corpus Directory:

Enter query keywords:   LSI

Lower Rank #:  Rank range is: {1,145}

Total no of documents: 145 Total no of words: 2541

Keyword List:

1: Georeferenced Population Data Sets of Mexico (GEO-MEX): Population Database of Mexicoxxxxxx (exclude)  
2: U.S. Population Grids (Summary File 1), 2000: Alabama, Louisiana, Mississippi and Texas, Alpha Versionxxxxxx (exclude)  
3: U.S. Population Grids (Summary File 1), 2000: New Orleans Metropolitan Statistical Area, Alpha Versionxxxxxx (exclude)  
4: Georeferenced Population Data Sets of Mexico (GEO-MEX): Urban Place Time-Series Population of Mexicoxxxxxx (exclude)  
5: U.S. Population Grids (Summary File 1), 2000: Houston Metropolitan Statistical Area, Alpha Versionxxxxxx (exclude)  
6: Georeferenced Population Data Sets of Mexico (GEO-MEX): Raster Based GIS Coverage of Mexican Populationxxxxxx (exclude)  
7: U.S. Population Grids (Summary File 3), 2000: New Orleans Metropolitan Statistical Area, Alpha Versionxxxxxx (exclude)  
8: U.S. Population Grids (Summary File 3), 2000: Alabama, Louisiana, and Mississippi, Alpha Versionxxxxxx (exclude)  
9: Georeferenced Population Data Sets of Mexico (GEO-MEX): Urban Place GIS Coverage of Mexicoxxxxxx (exclude)  
10: Gridded Population of the World, Version 1 (GPWv1)  
11: Georeferenced Population Data Sets of Mexico (GEO-MEX): GIS of Mexican States, Municipalities and Islandsxxxxxx (exclude)  
12: China Dimensions Data Collection: China County-Level Data on Population (Census) and Agriculture, Keyed to 1:1M GIS Mapxxxxxx (exclude)  
13: Global 15 x 15 Minute Grids of the Downscaled Population Based on the SRES B2 Scenario, 1990 and 2025  
14: Low Elevation Coast Zone (LECZ) Urban-Rural Population Estimates, Global Rural-Urban Mapping Project (GRUMP), Alpha Version  
15: Gridded Population of the World: Future Estimates (GPWFE)  
16: Gridded Population of the World, Version 3 (GPWv3)  
17: Gridded Population of the World, Version 2 (GPWv2)  
18: Country-Level Population and Downscaled Projections Based on the SRES A1, B1, and A2 Scenarios, 1990-2100  
19: Country-Level Population and Downscaled Projections Based on the SRES B2 Scenario, 1990-2100  
20: U.S. Census Grids (Summary File 1), 2000  
21: U.S. Census Grids (Summary File 1), 2000: Metropolitan Statistical Areas  
22: U.S. Census Grids (Summary File 3), 2000xxxxxx (exclude)  
23: U.S. Census Grids (Summary File 3), 2000: Metropolitan Statistical Areas

# Summary

- A data mining approach to improve geospatial data discovery
- LSA outperforms full-text search
- An alternative approach to establish domain ontology to complement the top-down approach

# Future Work

- Integration with Geonetwork
- Refine algorithm

# References

- W. Li, Automated Data Discovery, Reasoning and Ranking in Support of Building an Intelligent Geographic Search Engine, Ph.D. Dissertation. George Mason University, August 2010.
- W. Li, C. Yang, D. Nebert, R. Raskin and H. Wu, 2011. Semantic-based service chaining for building a virtual Arctic spatial data infrastructure. *Computers & Geosciences*. 37(11), 1752-1762.
- W. Li, R. Raskin and M.F. Goodchild, 2012a. Semantic similarity measurement based on knowledge mining: an artificial neural network approach. *International Journal of Geographic Information Science*, 26(8), 1415-1435.
- W. Li, M.F. Goodchild and R. Raskin, 2012b. Towards geospatial semantic search: exploiting latent semantic relations in geospatial data. *International Journal of Digital Earth*, DOI:10.1080/17538947.2012.674561.