



1st International KEYSTONE Conference
IKC 2015
Coimbra Portugal, 8-9 September, 2015

Keyword-Based Search over Environmental Datasets

José R.R. Viqueira

Alberto Bugarín

Joaquín Triñanes

Jaime Martínez-Urtaza



Centro Singular de Investigación
en Tecnoloxías da
Información





Outline

- ♦ **MOTIVATIONAL EXAMPLES**
- ♦ **ENVIRONMENTAL DATASETS**
- ♦ **KEYWORD-BASED SEARCH**
- ♦ **PROPOSED METHODOLOGY**



MOTIVATIONAL EXAMPLES

Motivational
Examples



♦ Public health and water microbiology (Cholera risk)

- High water temperatures and rainfall close to sea level during monsoon season

Environmental
Datasets

KeyWord-
Based Search

Proposed
Methodology

Keyword-Based Search over Environmental Datasets

♦ Public health and water microbiology (Cholera risk)

- High **water temperatures and rainfall** close to sea level during monsoon season

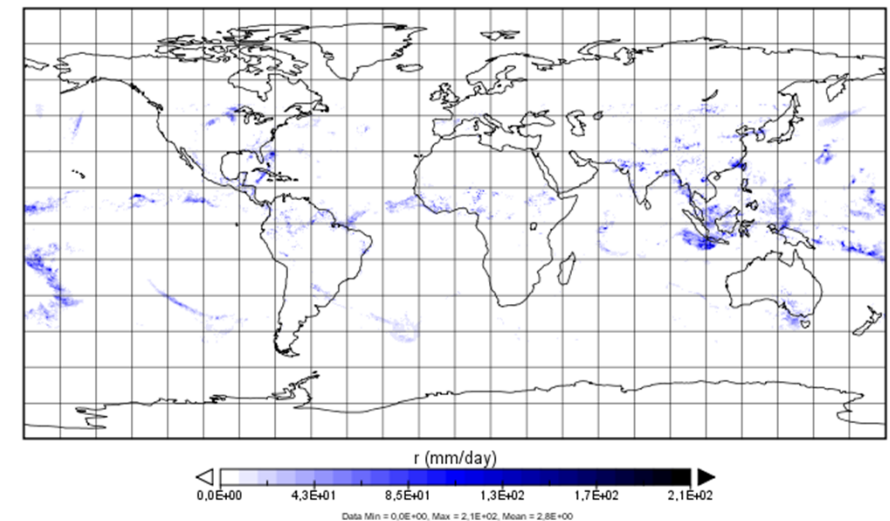
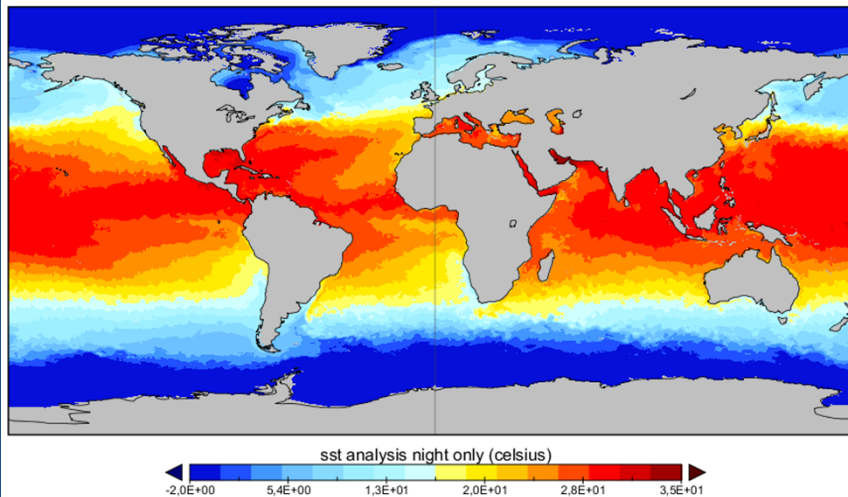
➤ Environmental properties

- **Sea Surface Temperature (SST)** [NOAA]
- **Rainfall (TRMM)** [NASA]

➤ Geographic Coverages (fields, grids, etc.)

- Very large scientific data multidimensional arrays
- Dims: Lat., Lon., Time, Depth, etc.

Many datasets
even for the
same property

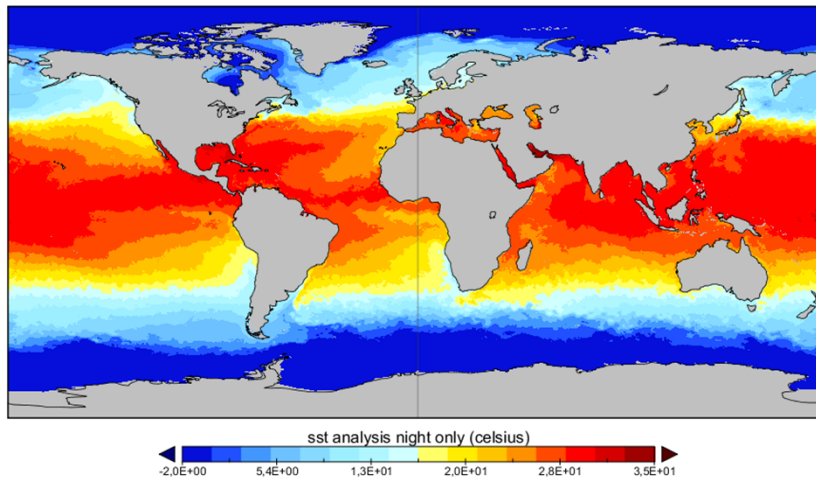


Public health and water microbiology (Cholera risk)

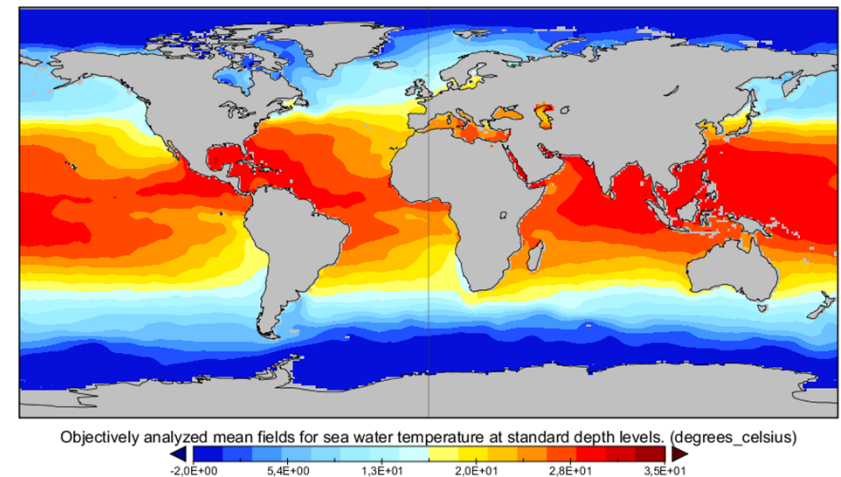
- **High** water temperatures and rainfall close to sea level during monsoon season

- Fuzzy linguistic value
- Relative to the mean of the same month during the last 7 years for the specific location
 - World Ocean Atlas (WOA). Statistical mean of temperature

Meaning depends on the context of application, geo location, time period, ...



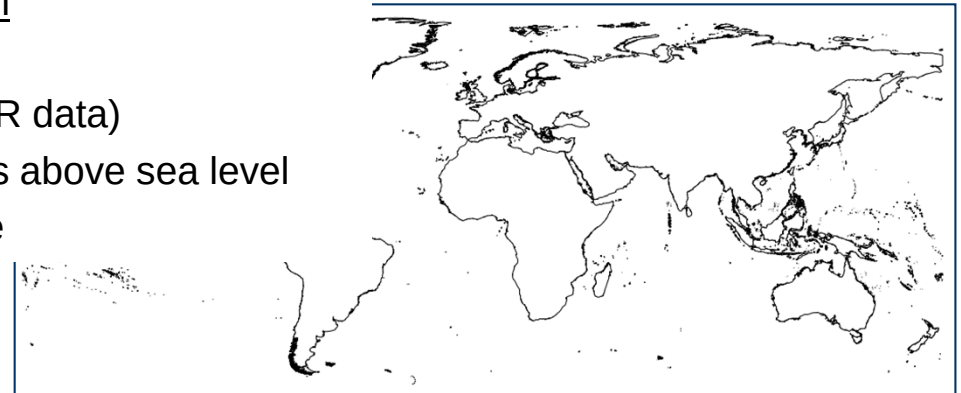
NOAA. cwblendednightSST. 22/07/2015



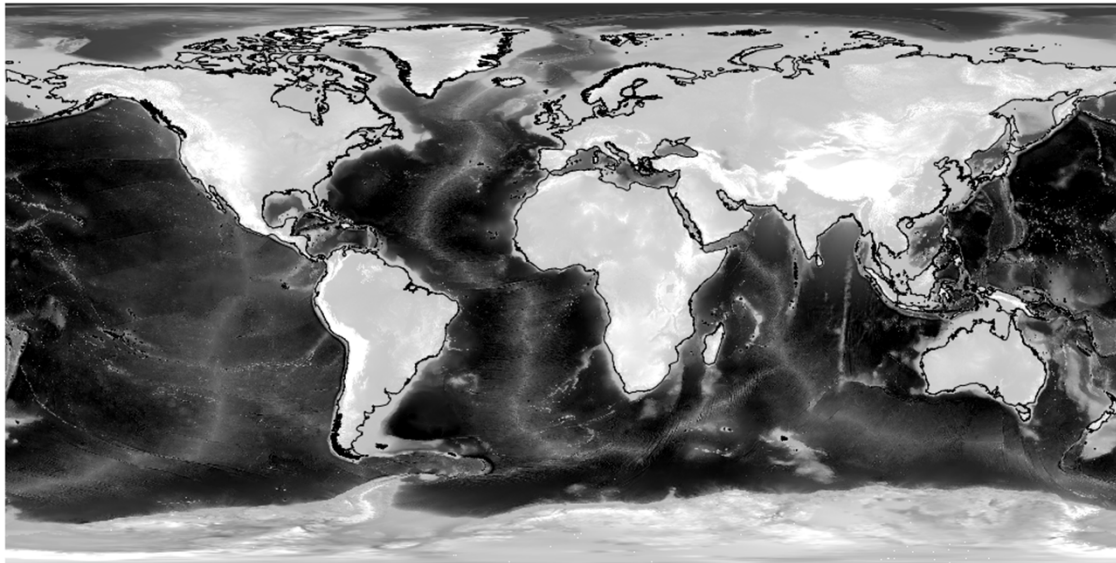
WOA. Stat. mean temperature. July

- ◆ **Public health and water microbiology (Cholera risk)**
 - High water temperatures and rainfall **close to sea level** during monsoon season

- Fuzzy geographic restriction
- Close to the **coast line**
 - Geographic Entity (E/R data)
- **Elevation** of about 0 meters above sea level
 - Geographic Coverage



Coast line. NOAA GSSHG



**Elevation.
NOAA ETOPO1**

Motivational
Examples



♦ Public health and water microbiology (Cholera risk)

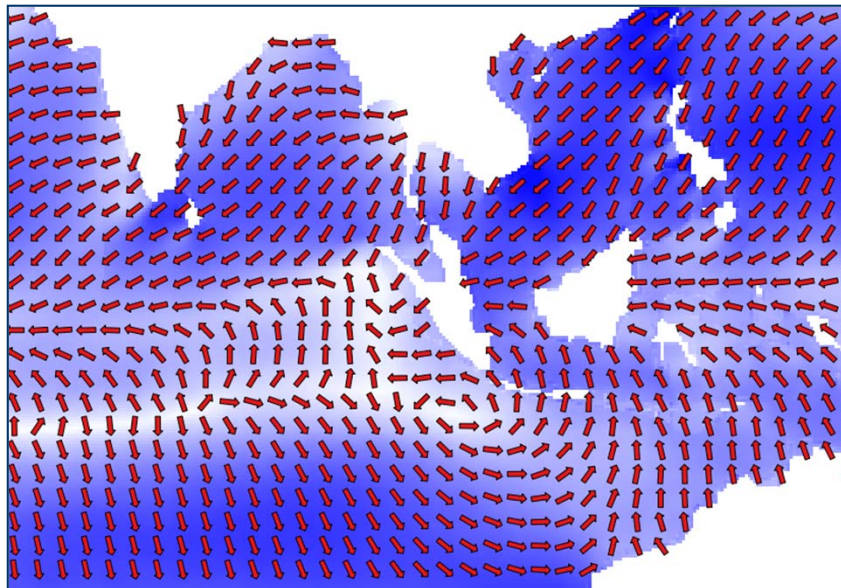
- High water temperatures and rainfall close to sea level **during monsoon season**

- Fuzzy temporal restriction
- Monsoon season (Malaysia)
 - Northeast Monsoon from November to March (more rainfall)
 - Southwest Monsoon from May to September (less rainfall)

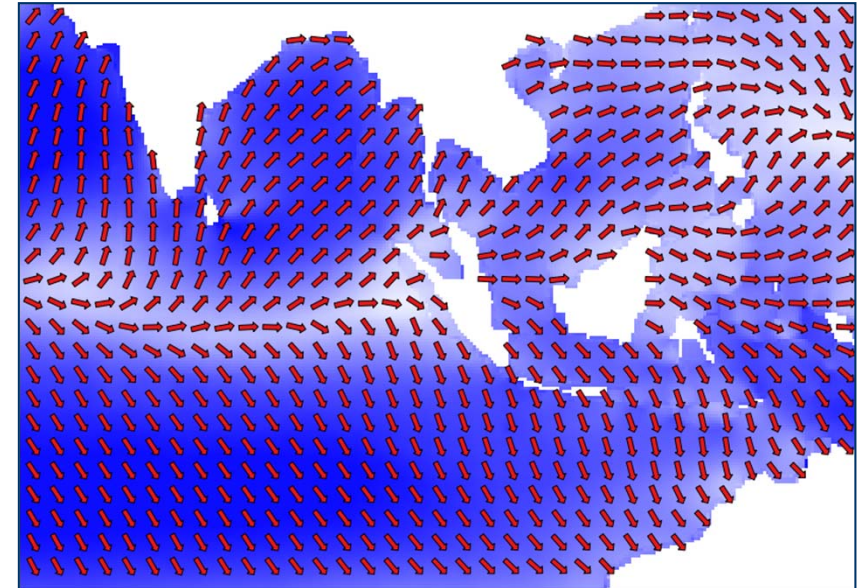
Environmental
Datasets

KeyWord-
Based Search

Proposed
Methodology



NOAA Blended Sea Winds. January



NOAA Blended Sea Winds. July



MOTIVATIONAL EXAMPLES

Motivational
Examples



Environmental
Datasets

KeyWord-
Based Search

Proposed
Methodology

♦ Severe weather reporting

- Hurricane of high category close to highly populated areas
- Strong winds and heavy swell on marine traffic routes
- Hailstorm on the highway

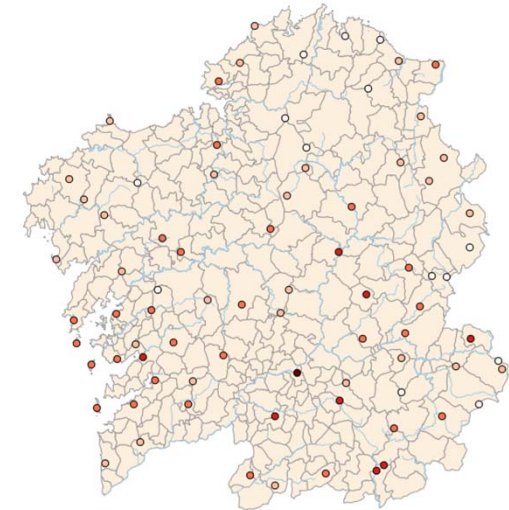
♦ Tourism

- Beach protected from wind, with warm water and air temperature, and few waves

◆ Do not always fit E/R model

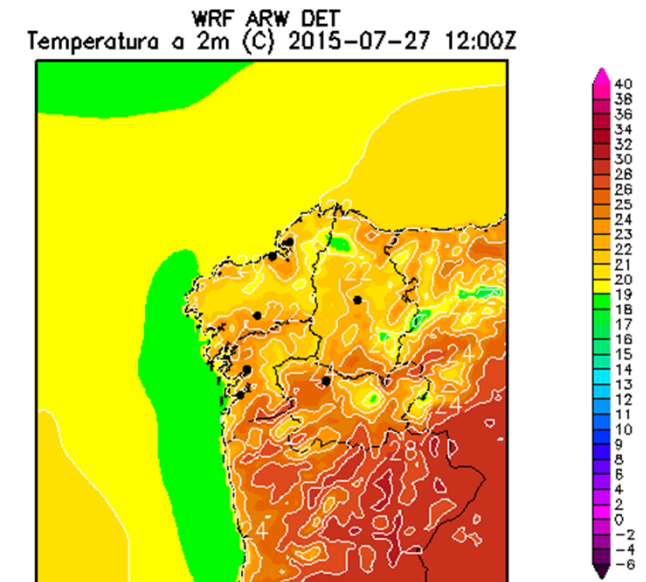
■ Geographic Entities

- Entities with geometric properties
- Relational data
- Rivers, Meteorological stations, Municipalities, etc.



■ Geographic Coverages

- Mappings with geographic domain
- Multidimensional array data
- Temperature, rainfall, elevation, salinity, etc.



MeteoGalicia - CMAT - Xunta de Galicia

Motivational
Examples

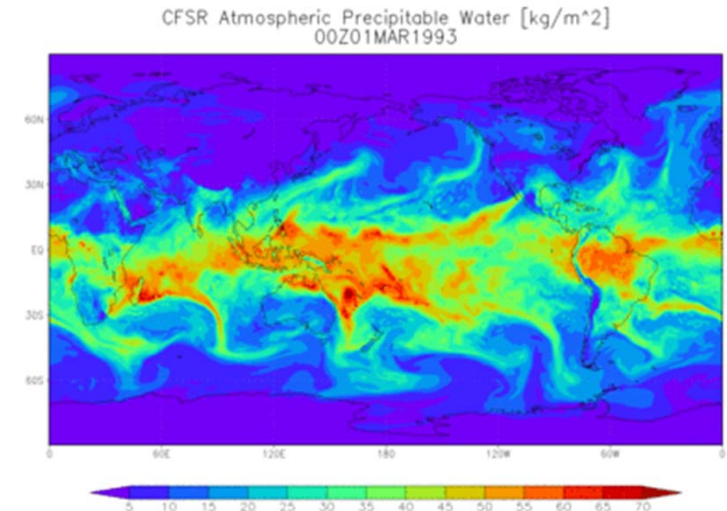
Environmental
Datasets

Keyword-
Based Search

Proposed
Methodology

♦ Many large or very large datasets

- Stations, buoys, satellites, radars, etc.
 - Blended Sea Winds
 - Over 300 gigabytes
- Prediction, reanalysis, etc.
 - Climate Forecast System Version 2 (CFSv2)
 - Over 500 terabytes



Motivational
Examples

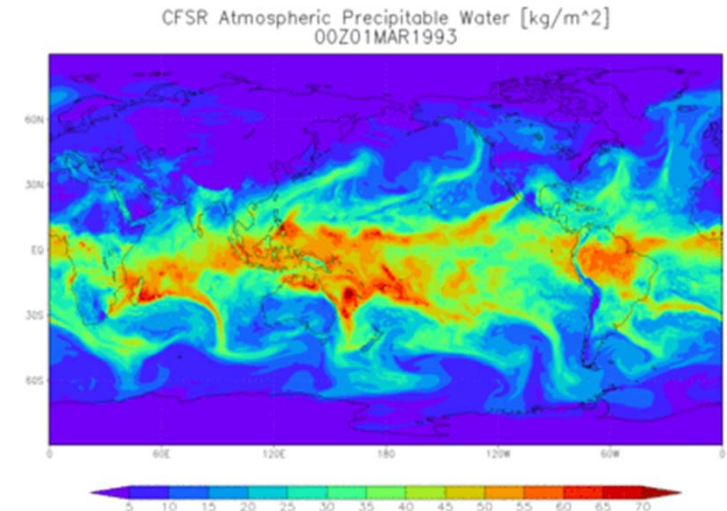
Environmental
Datasets

Keyword-
Based Search

Proposed
Methodology

♦ Many large or very large datasets

- Stations, buoys, satellites, radars, etc.
 - Blended Sea Winds
 - Over 300 gigabytes
- Prediction, reanalysis, etc.
 - Climate Forecast System Version 2 (CFSv2)
 - Over 500 terabytes



♦ Highly heterogeneous

- Formats and encodings
 - Shape Files, GML, GeoJSON, etc.
 - GeoTiff, ASC, NetCDF, HDF, GRIB, etc.
- Semantics
 - Temperature, Sea Temperature, Sea Surface Temperature, air temperature, etc.

Motivational
Examples

Environmental
Datasets



Keyword-
Based Search

Proposed
Methodology

♦ Open Data

- Spatial Data Infrastructures (SDIs)
 - Global Spatial Data Infrastructure (GSDI)
 - INSPIRE
 - NSDI
- Unidata
- OpenDAP





ENVIRONMENTAL DATASETS

Motivational
Examples

Environmental
Datasets



Keyword-
Based Search

Proposed
Methodology

♦ Open Data

- Spatial Data Infrastructures (SDIs)
 - Global Spatial Data Infrastructure (GSDI)
 - INSPIRE
 - NSDI
- Unidata
- OpenDAP



♦ Data models, Vocabularies, Ontologies

- Observations and Measurements (O&M)
- Sensor Model Language (SensorML)
- Semantic Web for Earth and Environmental Terminology (SWEET)
- CF Conventions and Metadata
- World Meteorological Organization Manual on Codes
- SeaDataNet Sensor Vocabularies
- ETC.



KEYWORD-BASED SEARCH

Motivational
Examples

Environmental
Datasets

KeyWord-
Based Search



Proposed
Methodology

♦ Non structured data sources

- Information Retrieval (IR)
 - Text data
- Multimedia Information Retrieval (MIR)
 - Audio, images, video.

Google

bing

Y!
YAHOO!

Motivational
Examples

Environmental
Datasets

KeyWord-
Based Search

Proposed
Methodology

◆ Non structured data sources

- Information Retrieval (IR)
 - Text data
- Multimedia Information Retrieval (MIR)
 - Audio, images, video.

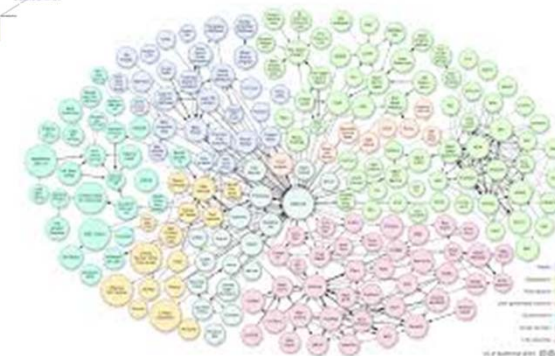
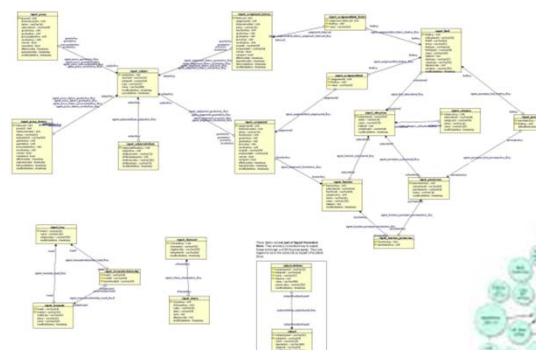
Google

bing

Y!
YAHOO!

◆ Structured data sources

- Relational DBMSs
- Linked Data
- **E/R based data!**



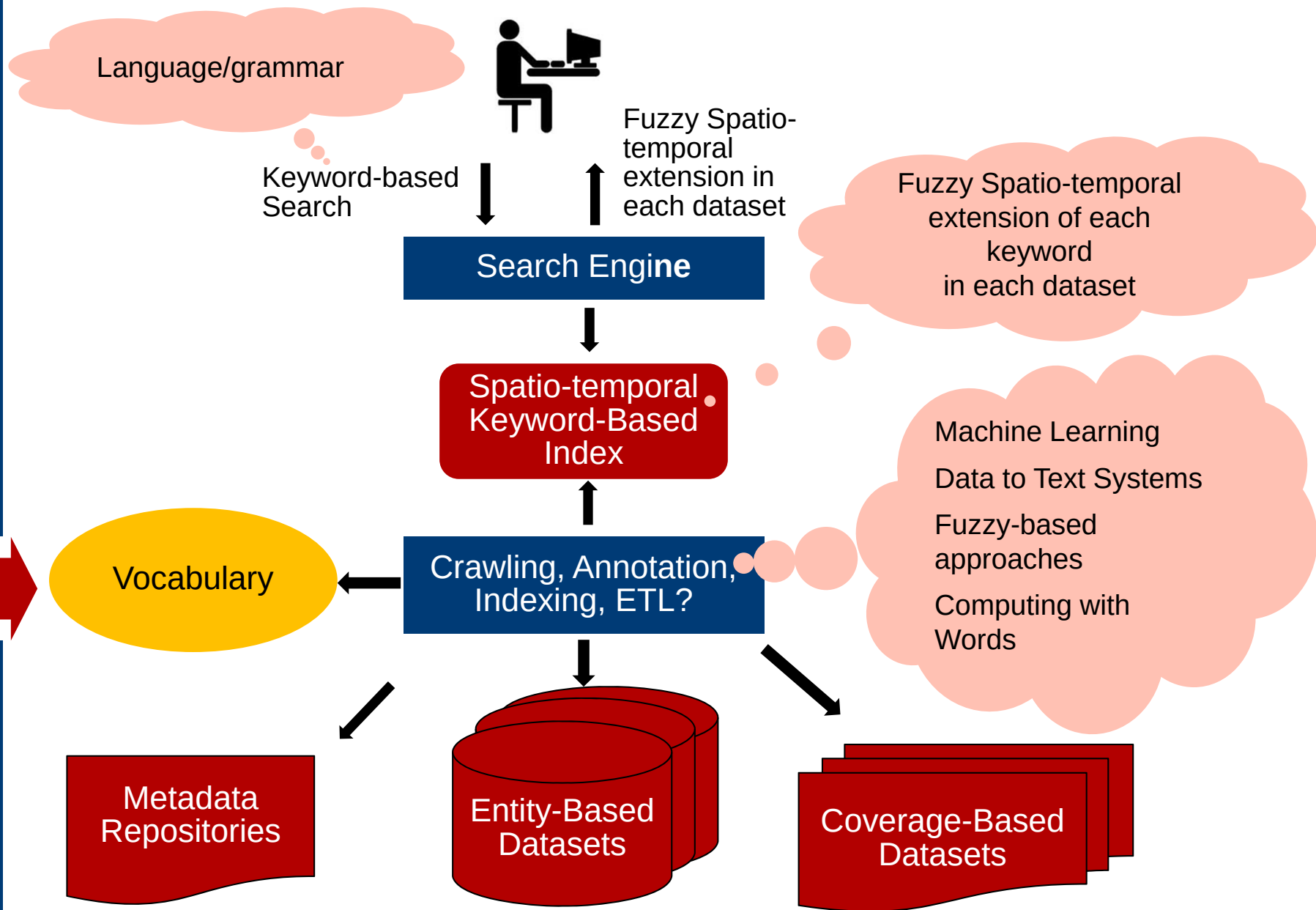
PROPOSED METHODOLOGY

Motivational
Examples

Environmental
Datasets

KeyWord-
Based Search

**Proposed
Methodology**



Keyword-Based Search over Environmental Datasets



1st International KEYSTONE Conference
IKC 2015
Coimbra Portugal, 8-9 September, 2015

Keyword-Based Search over Environmental Datasets

José R.R. Viqueira (jrr.viqueira@usc.es)

Alberto Bugarín (alberto.bugarin@usc.es)

Joaquín Triñanes (joaquin.trinanes@usc.es)

Jaime Martínez-Urtaza (j.l.martinez-urtaza@bath.ac.uk)



Centro Singular de Investigación
en Tecnoloxías da
Información



UNIVERSITY OF
BATH