Sabin Center for Climate Change Law   Research Centers & Programs

2016

EIS Database Design Considerations

Sritharan Vinayagamoorthy

Center for International Earth Science Information Network

EARTH INSTITUTE | COLUMBIA UNIVERSITY

EIS Database Design Considerations

Sri Vinay

Associate Director for IT
CIESIN, Columbia University

16 June 2016

EIS Database Design Considerations

Objective and Scope

- Focus on regional and local (city etc.) level EIS of interest in addition to what is available at:
    - Northwestern Transportation Library (32,000+)
    - HathiTrust (33,000+)
    - EPA (14,460)
    - DOE (425)
    - Others?

- Add new capabilities and support new use cases that are complementary
    - Georeference the statements, whenever possible, and enable GIS capabilities.
    - Identify and link related statements?
    - Additional information or data such as imageries or spatial boundaries?

- Architecture and design a sustainable system
    - For statements already available on-line, store only the necessary metadata but leave the statements in their original location.
    - Statements that are not on-line but available for ingest, we could potentially consider archiving them.

2

EIS Database Design Considerations

Metadata

- Use advanced techniques such as Natural Language Processing (NLP) and Machine Learning Algorithms (MLA) to automatically parse and extract useful metadata from the statements

    - Spatial information (coordinates) for georeferencing the statements
    - Title
    - Author
    - Subject
    - Abstract
    - Temporal information (time period)
    - Publisher
    - Date of publication
    - Language
    - Format
    - Maps/Pictures?
    - Charts/Graphs
    - Data tables?
    - ??

EIS Database Design Considerations

Capabilities (Should be driven by Use Cases)

- Search

  - Full text and/or any of the metadata fields listed in previous slide
  - Logical (AND/OR) combination of search criteria
  - Map driven, geospatial search by bounding box, polygon, and shape file
  - Geospatial search by place name (using a gazetteer service)
  - Semantic search (ontology; pH/acidity/alkalinity)
  - Drill down by metadata facets
  - ??

- Presentation

  - Display results in an interactive map?
  - Show related imageries and additional information if available?
  - Provide maps, pictures, charts, and data tables if available
  - Sorting capabilities?
  - Grouping capabilities (related statements)?
  - ??

4

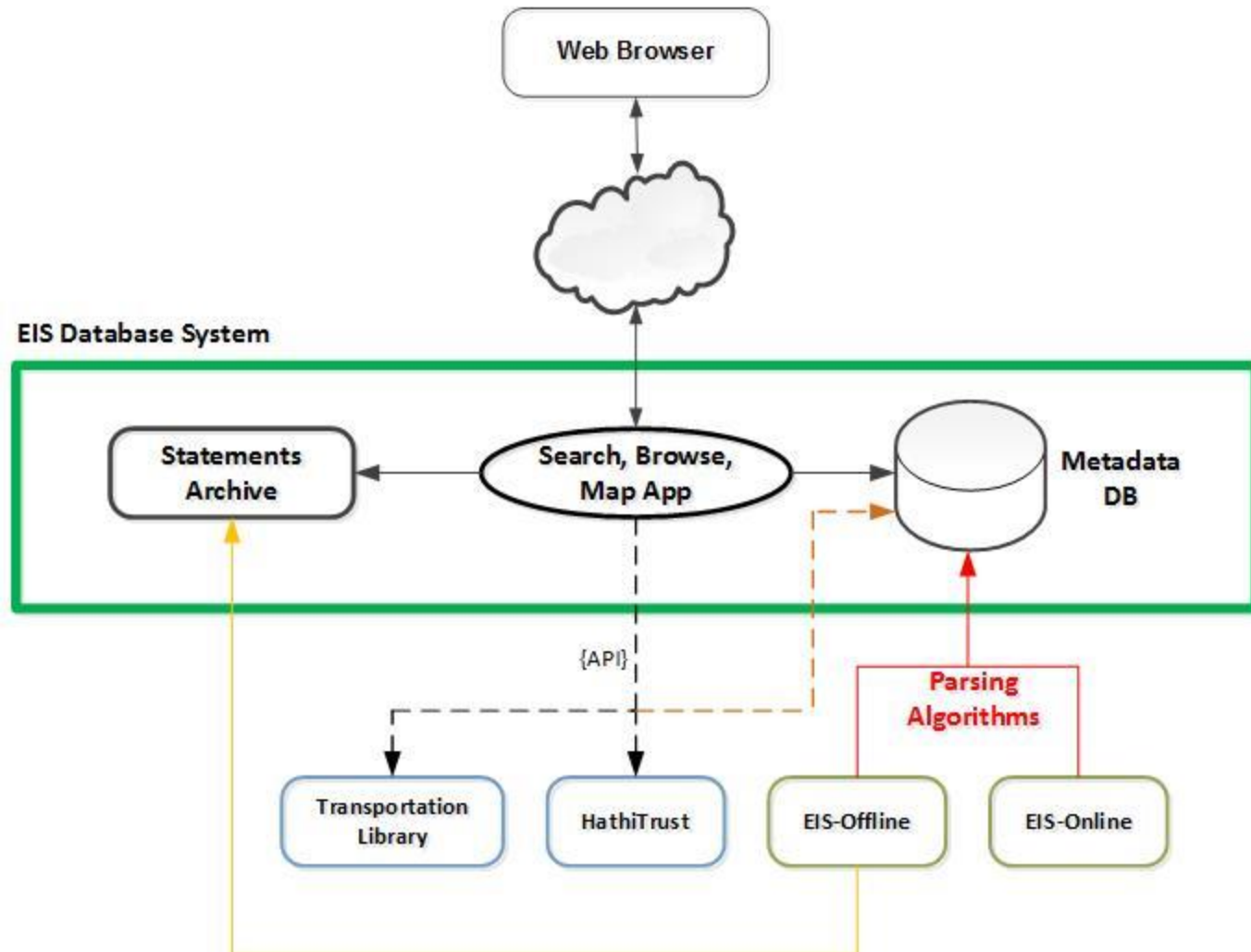# EIS Database Design Considerations

Interoperability

- **External Resources**

    - Access external EIS systems (e.g. HathiTrust) through web services and/or APIs if available
    - Access cloud hosted base maps, imageries, etc. via web services
    - Access external Gazetteer services

- **Internal Resources**

    - Make query, results, and metadata available via web services
    - REST API (XML / JSON)
    - OGC WMS / WFS services for georeferenced statements

EIS Database Design Considerations

Architecture

EIS Database Design Considerations

Technology

- Open Source

  - PostgreSQL/PostGIS (Spatial-Relational DB)
  - Apache Solr (Search)
  - Geoserver (OGC Mapping services)
  - Tomcat/Java (REST services)
  - OpenLayers/Leaflet (Mapping/Visualization Tools)
  - Fedora (Archiving)

- COTS

  - ESRI ArcGIS (Spatial)
  - Oracle (Relational DB)
  - Google Search

- Hosting Options

  - Internal Cloud (VMware based virtual server environment)
  - External Cloud (Amazon etc.)