Components of a Scaleable Geo ICT Compute Infrastructure

Profs. N.L. Sarada & Umesh Bellur

Geo spatial Information Science and Engg. Lab

Dept. of CSE, IIT Bombay

Outline

- GISE Lab @ CSE, IIT Bombay
  - Structuring our research
- Our vision for a nation wide systems architecture for geo spatial data and applications
- Current status and some accomplishments

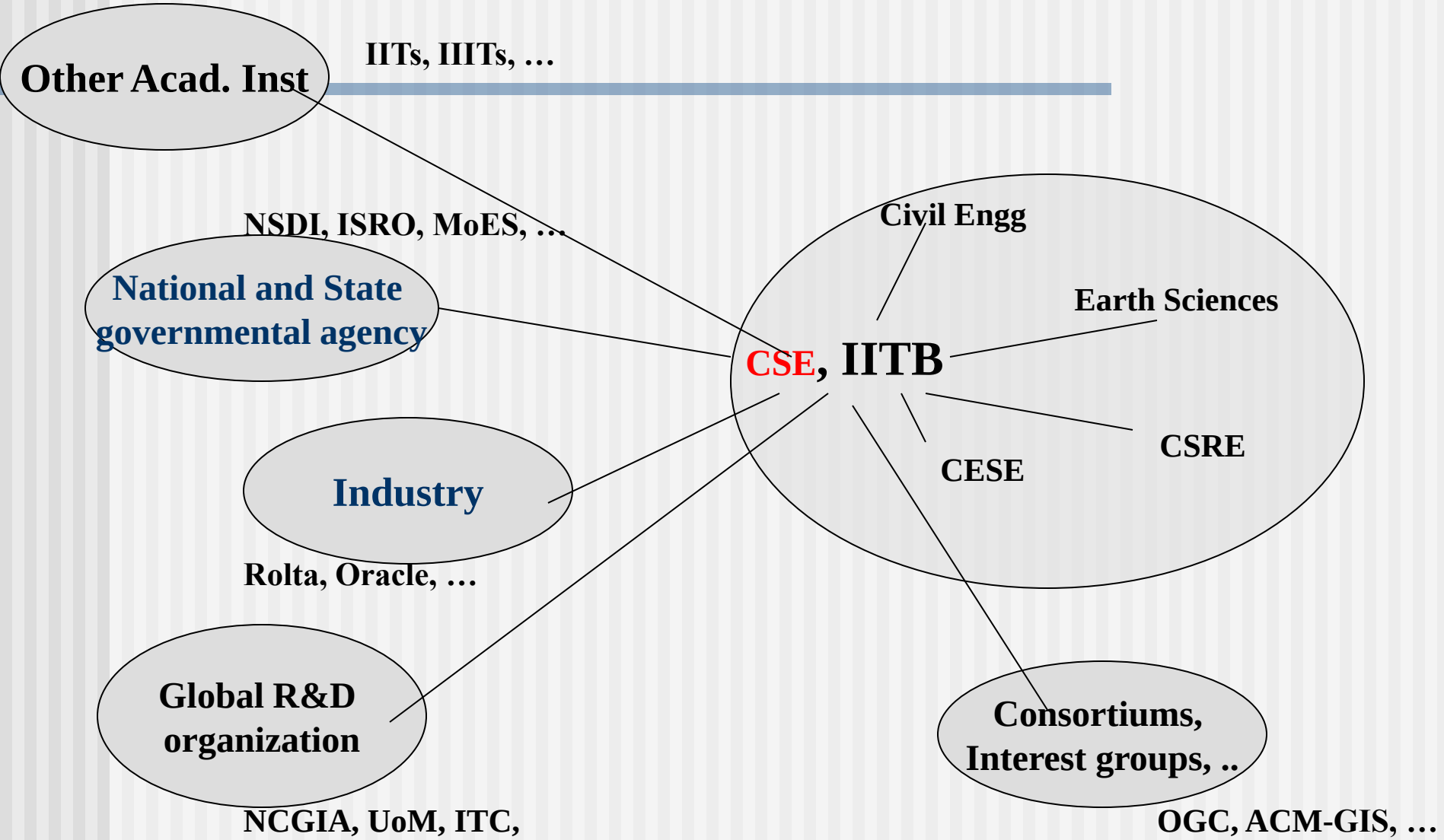
Outline

- GISE Lab @ CSE, IIT Bombay
  - Structuring our research
- Our vision for a nation wide systems architecture for geo spatial data and applications
- Current status and some accomplishments

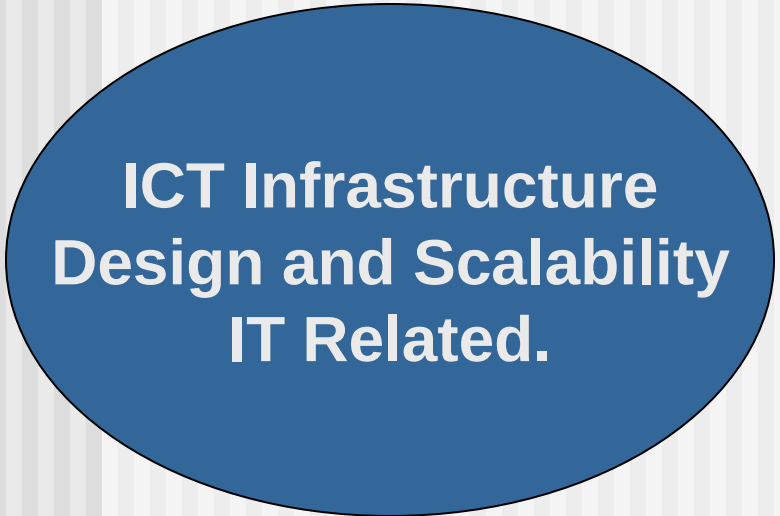
GISE – Scope and Objectives

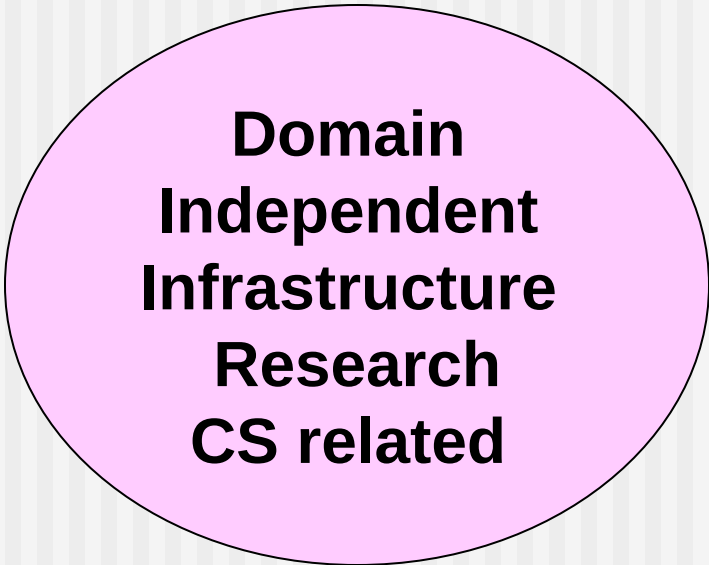
- Research focus of core technology areas for ***processing*** geo-spatial data
  - Computer science perspective
- Build research repository and technology showcase for the community
- Capacity building in research
  - Workshops and short-term courses
  - Research and graduate students

A collaboration centric approach

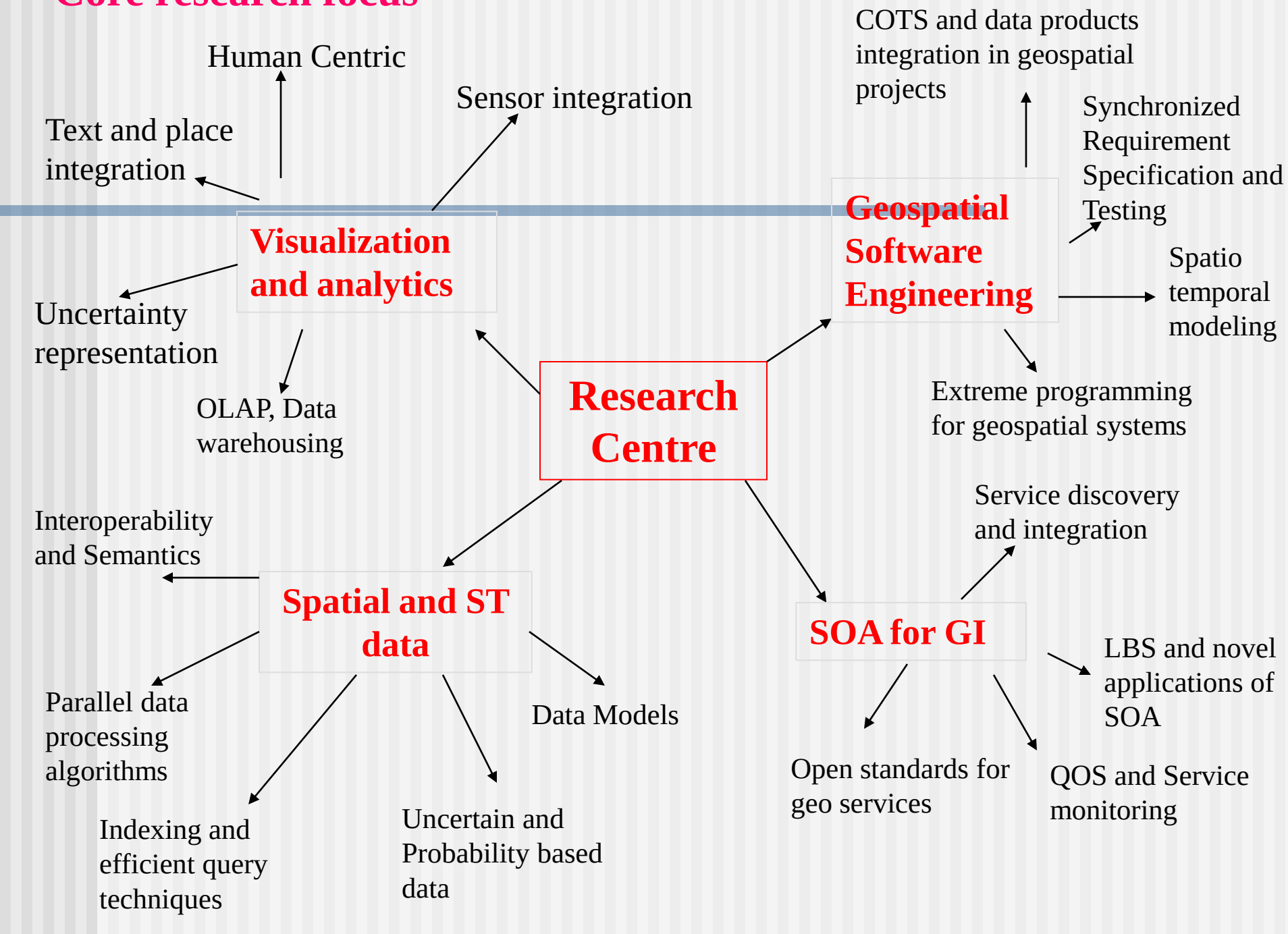


Carving up the Research Space





Core research focus



Infrastructure Research

Data Analytics –
Real time and Warehouse based mining
Geo Informatics

Model Search and Equivalence
Ontology Mapping & Equivalence

Data Collection Strategies – Sensor Networks, data aggregation patterns

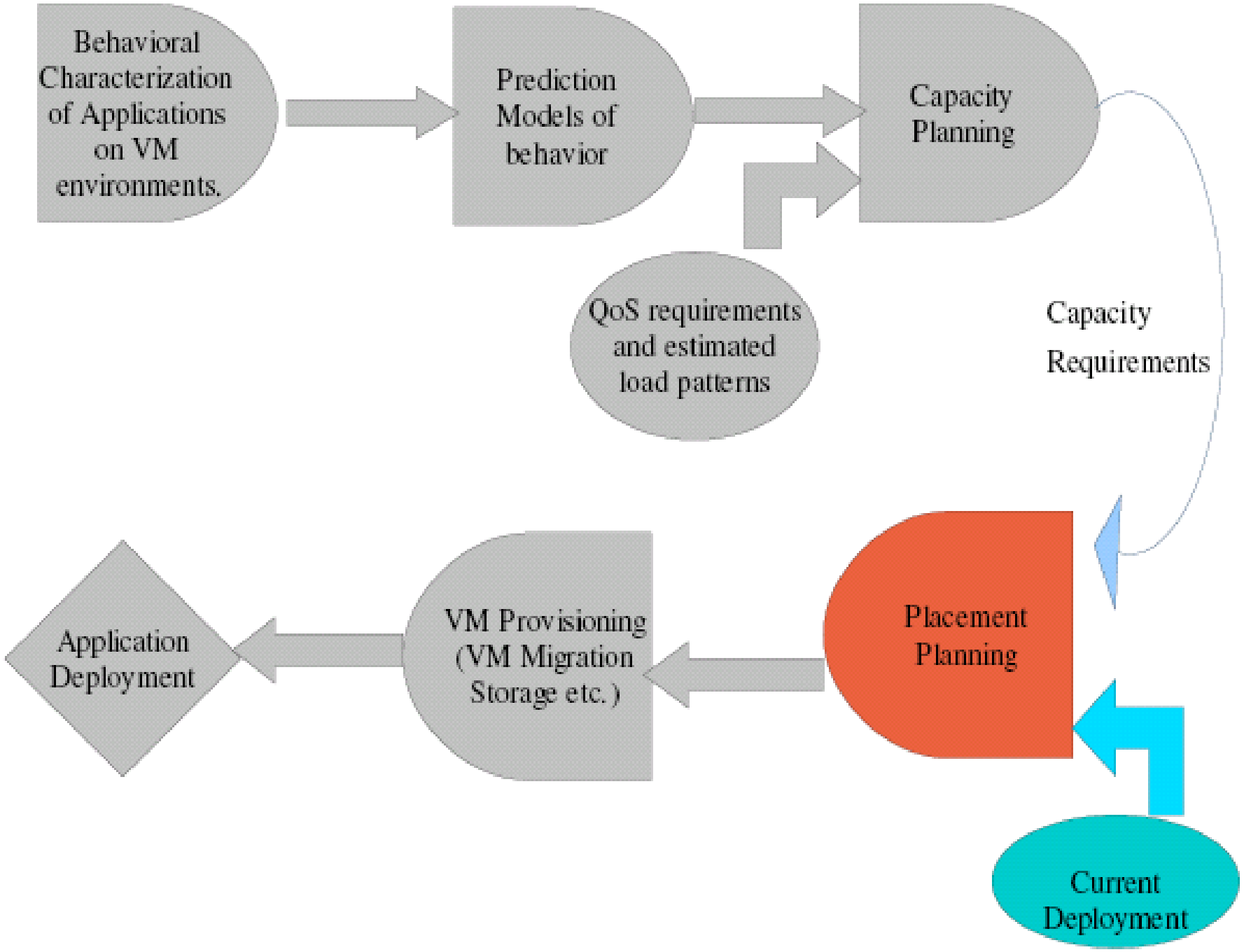
Data Model Repository Structure and mgmt

Meta Models and Ontology Creation Tools

STANDARDS Devl

IT related research

- Data center & Network design – use of virtualization and cloud computing
- Capacity planning
- Monitoring, alerts, autonomic computing

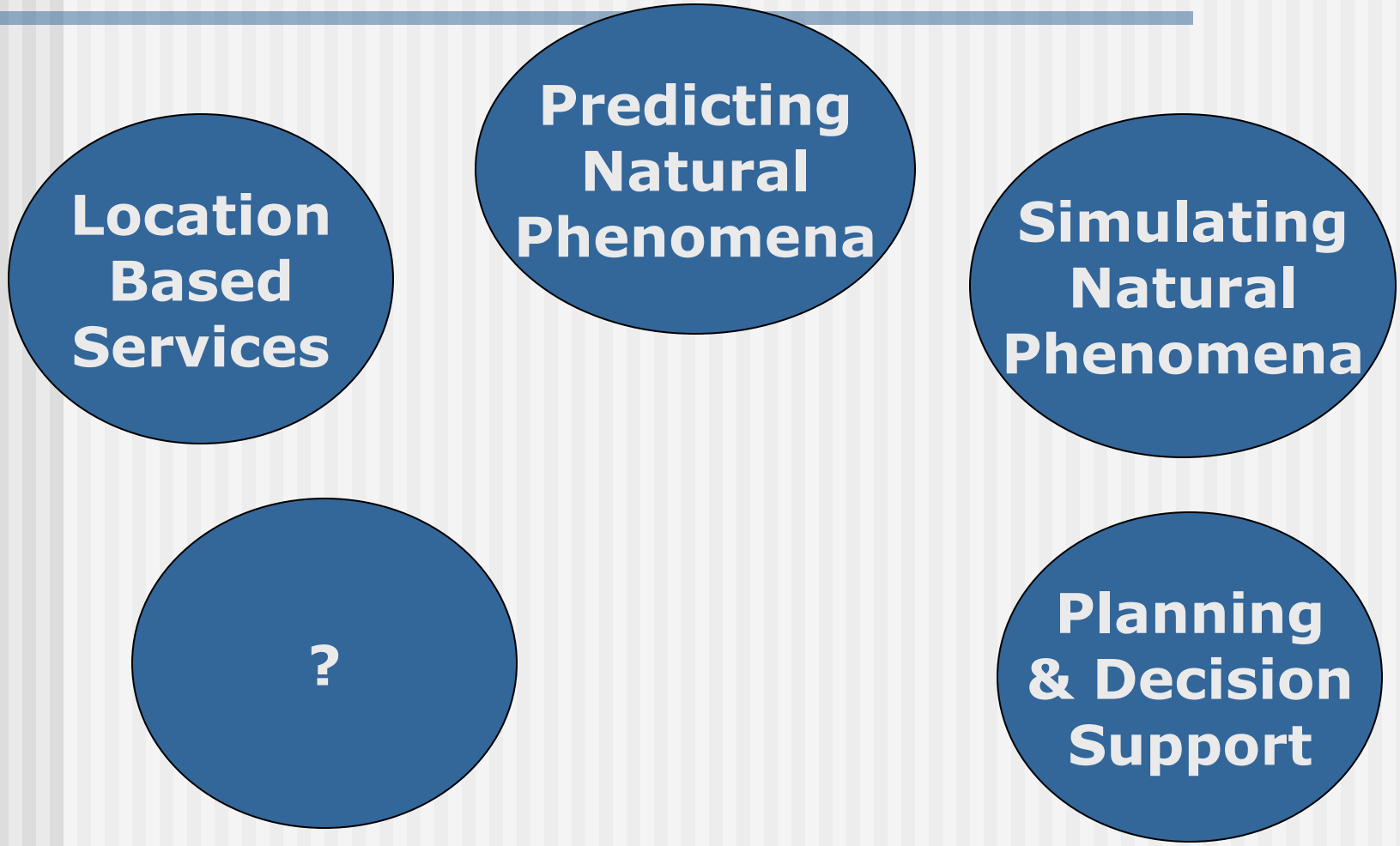

Behavioral Characterization of Applications on VM environments.
Prediction Models of behavior
Capacity Planning
QoS requirements and estimated load patterns
Capacity Requirements
Placement Planning
VM Provisioning (VM Migration Storage etc.)
Application Deployment
Current Deployment

What is Applications R&D?

Domain Specific Data Models

Domain Specific Ontologies – Water Mgmt, etc.

Some Application Areas



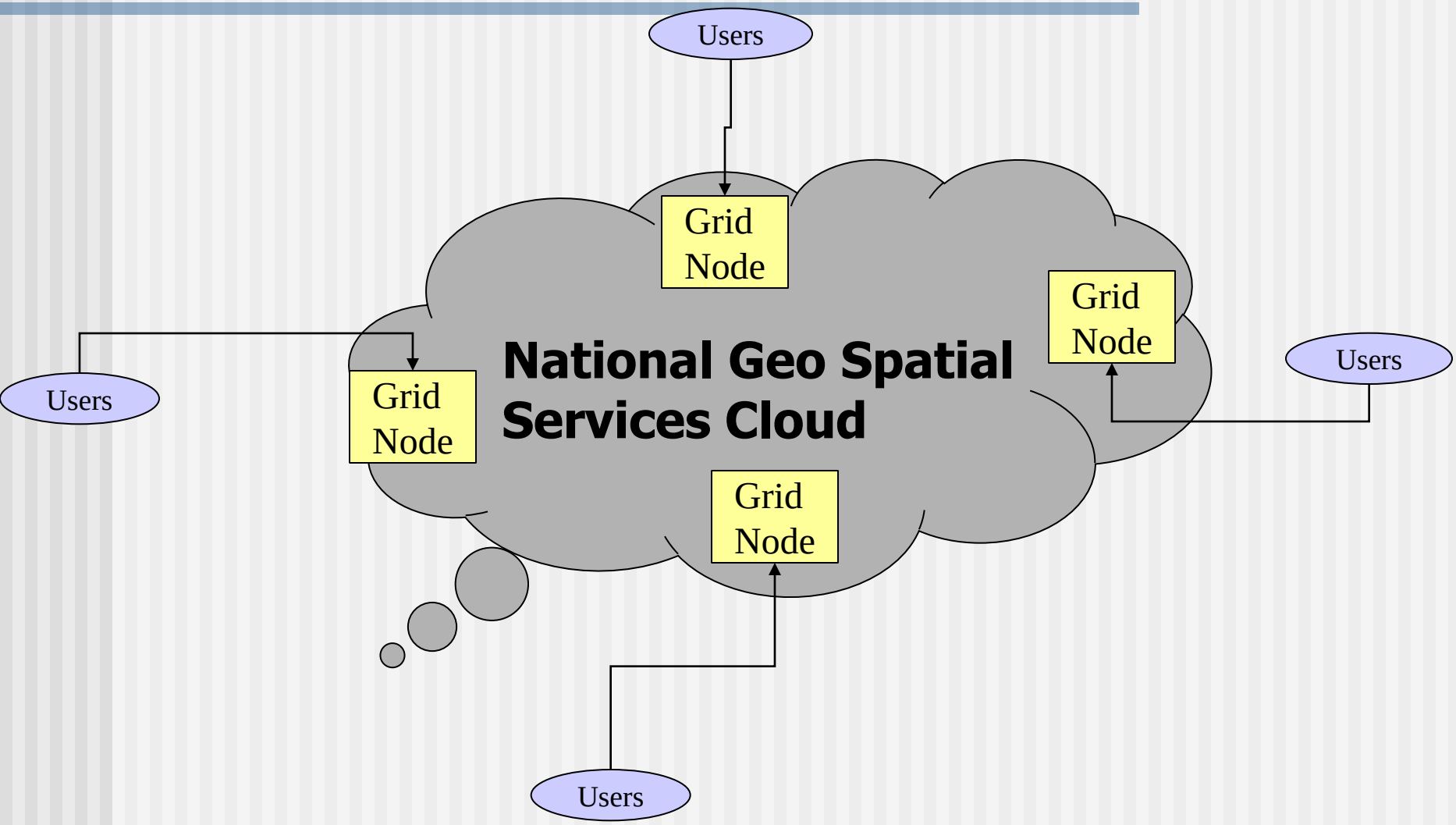
Outline

- GISE Lab @ CSE, IIT Bombay
  - Structuring our research
- Our vision for a nation wide systems architecture for geo spatial data and applications
- Current status and some accomplishments

Guiding principles

- Emulate a grid
  - Hardware can grow organically but is unified by software layers above.
- Federated (logically centralized) data availability but protected by layered, role based access control
- Web enable all services
- Be OGC standards compliant

The Vision – a National Geo Spatial data/services Grid



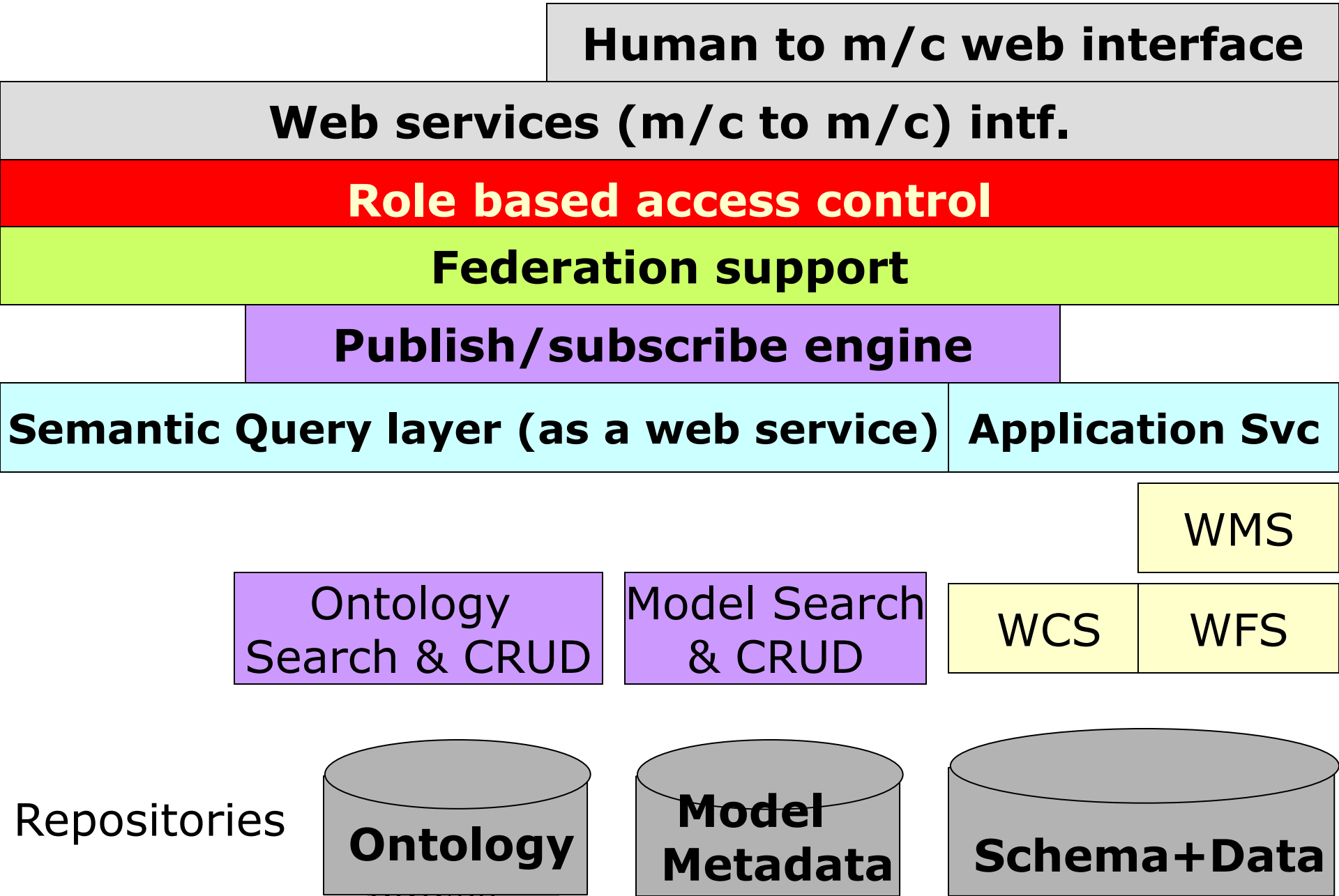
Requirements on a Grid Node

- Serve as a federated data store for natural resources data
  - Based on pre-defined, standardized data models
- Serve as a federated services repository of analytical applications.
  - Standardized ontologies needed

Requirements on the Grid

- Cooperation of nodes:
  - Services search
  - Data search
- Goal is to treat the set of all grid nodes as a single virtual node that hosts data and analytic service applications

GRID NODE ARCHITECTURE



Implications

- Network accessible instead of desktop bound.
  - Will enable a national geo spatial applications grid to be set up with a few data centres
- Standards Based
  - Web services, GML etc.
- Scalable backend
  - Oracle 10G, clustered application servers etc.

Outline

- GISE Lab @ CSE, IIT Bombay
  - Structuring our research
- Our vision for a nation wide systems architecture for geo spatial data and applications
- Current status and some accomplishments

Accomplishments

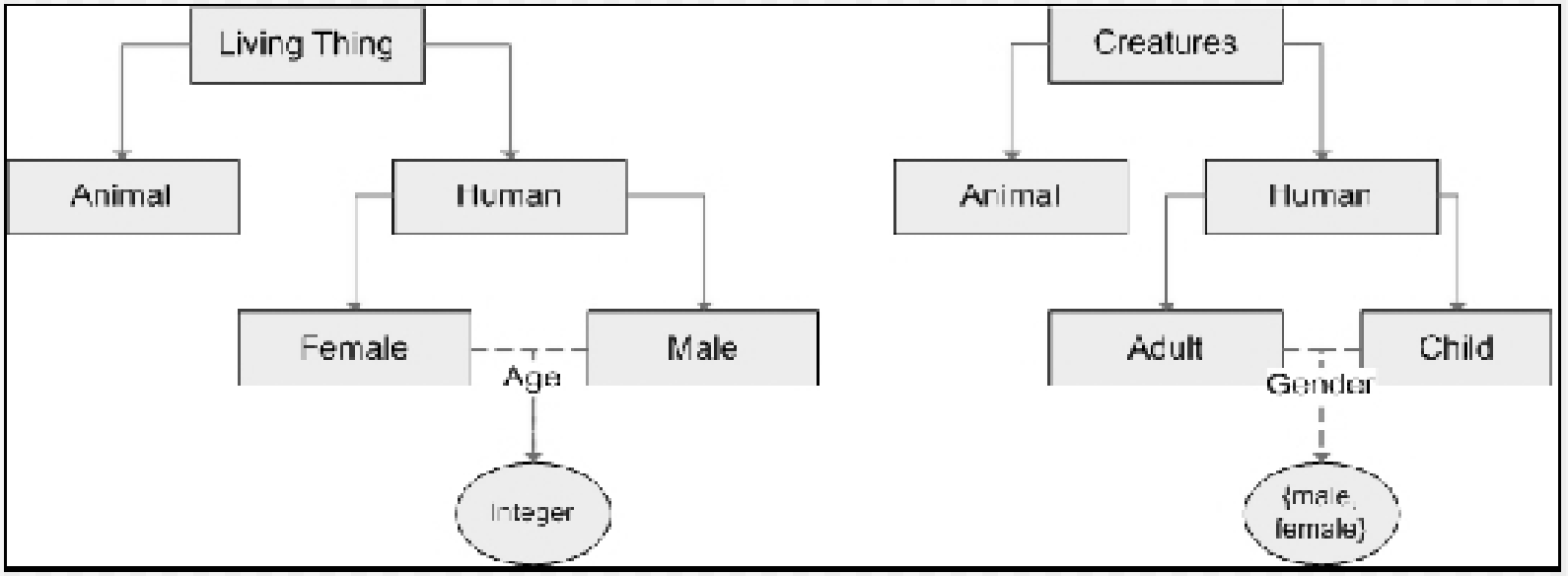
- Semantic Interoperability toolset
- Schema generator framework

Semantic Interoperability

- Multiple data sources each with it's own description model
- OGC and other standards addresses only *syntactic interoperability*
  - GML
  - WGS84 etc.
- But this does NOT address issues of semantic incompatibilities in different data sets.

The Semantics Problem

- Different entities names, same meaning.
- Same entity name, different meaning.
- Different classification hierarchies.
- Different attributes, units, …



Ontologies

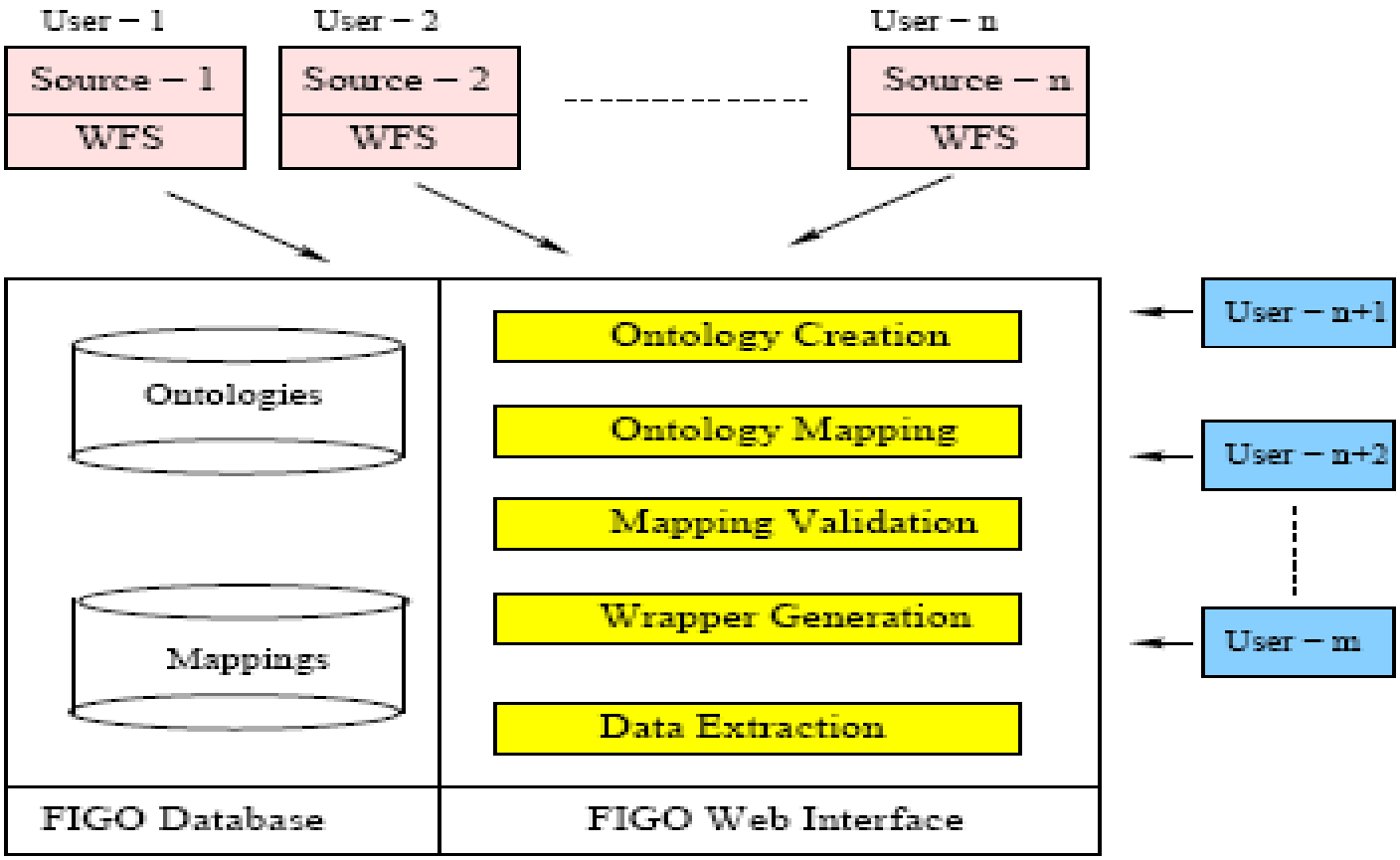
- An explicit specification of a shared conceptualization.
- Defined using the concepts of
  - Classes
  - Data and Object Properties
  - Constraints and Axioms
- Describes the semantics of the information sources, makes the content explicit.
- Can be used for the identification and association of semantically corresponding information concepts.

Our Solution - FIGO

Framework for Interoperability in GIS using Ontologies

- Use Ontologies to capture the meaning (**Meta data Entry)**
- Identify and associate semantically equivalent concepts across sources **(Ontology Mapping)**
- Check consistency of the mappings using ontology reasoners. **(Mapping Validation).**
- Accept data requests, transform them according to contents of sources.
- Extract data from sources and merge results

FIGO Architecture

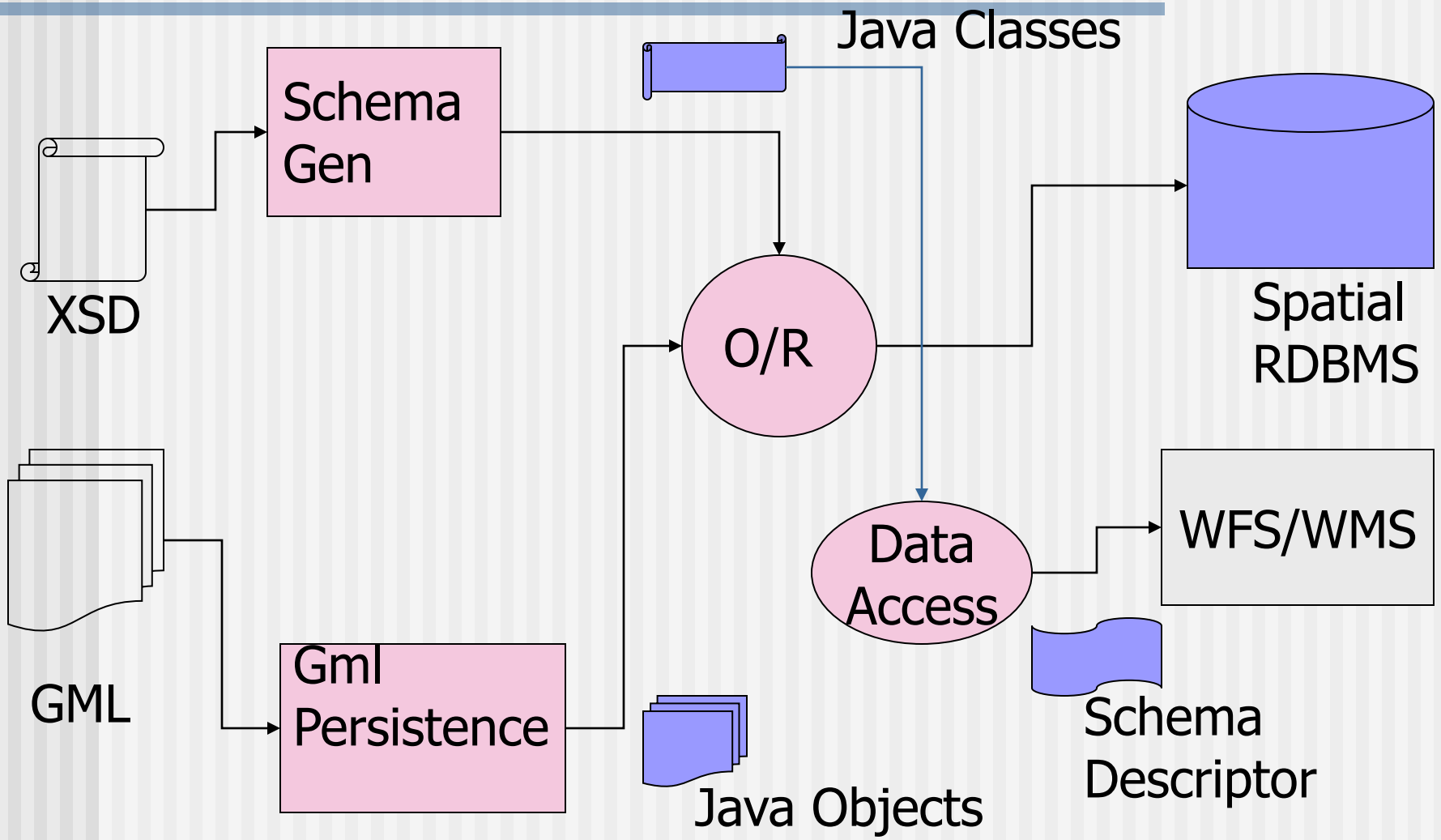


What is SGF?

- A collection of tools to:

1. Go from XML Schema Descriptions of geo spatial data to a relational schema
2. Insert GML data into the spatial schema so created.
3. Extract the data from spatial DBs using OGC standard services (WFS/WMS)

SGF Architecture



SGF Pros and Cons

- Pros:
  - Automates relational schema design and installation
  - Inserts data gotten as GML
  - Generates a corresponding Degeree descriptor to match generated schema.
- Cons:
  - Needs a XSD to begin with (currently)
  - Tied to Deegree - descriptor could be different for different WFS implementations.

Summarizing

- A comprehensive strategy is needed to collect, clean and make available data and applications to the public ***so that redundancy is avoided***
- Educating state SDIs and other organizations about this strategy and operational guidelines is critical for it's success.
- CS Research and progress in geo spatial technologies are not mutually exclusive

Questions?

GISE, Dept. of CSE, IIT Bombay

Prof. N.L. Sarada, Prof. Umesh Bellur

nls@cse.iitb.ac.in, umesh@cse.iitb.ac.in