Semantic Computing for Remote Sensing Big Data

Challenge and Opportunities

Lizhe Wang, Prof. Dr. –ing

Institute of Remote Sensing and Digital Earth
Chinese Academy of Sciences

Outline

1. Introduction and Background
2. Remote Sensing Big Data Semantic Organization
3. Remote Sensing Big Data Information Retrieval
4. Remote Sensing Big Data Semantic Service
5. Summary

1. Introduction and Background

☐ What is Semantic Computing

- Semantic computing is a field of computing that combines elements of semantic analysis, natural language processing, data mining and related fields.
- Semantic computing addresses three core problems:

✓ Understanding the intentions (semantics) of users and expressing them in a machine-processable format.

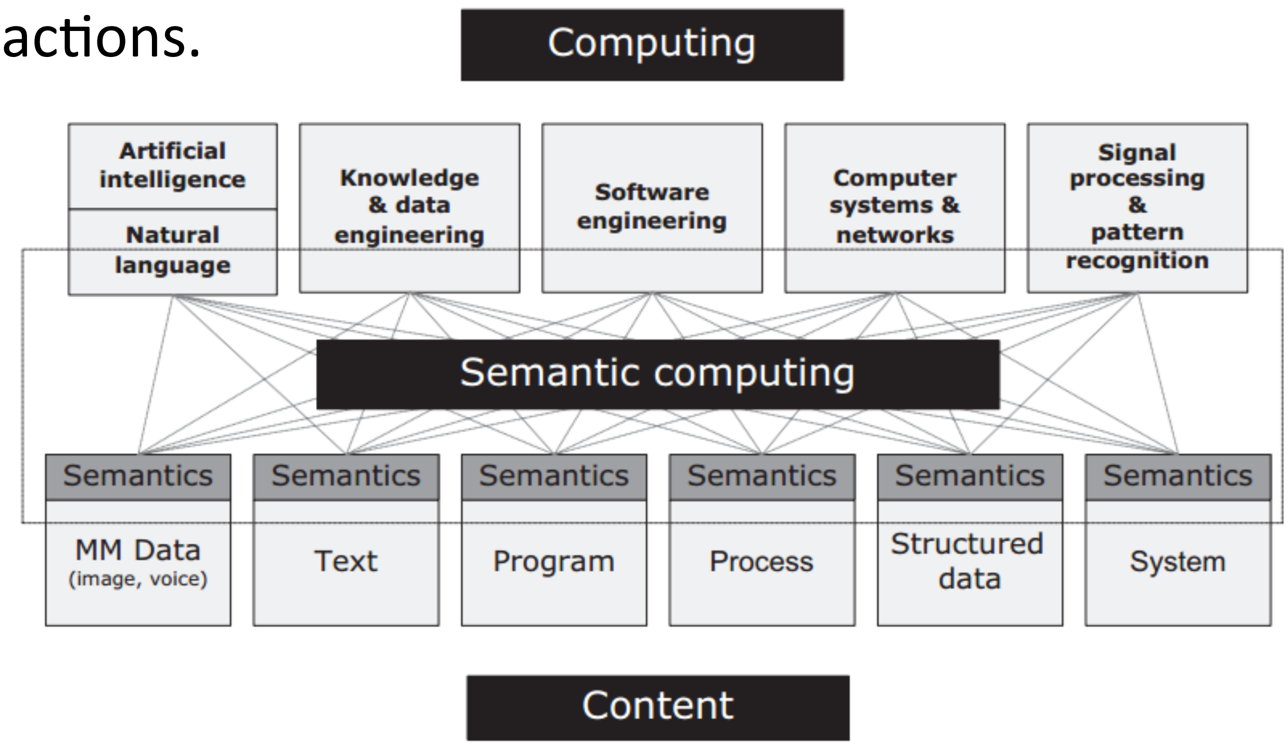
✓ Understanding the meanings (semantics) of computational content (of various sorts) and expressing them in a machine-processable format.

✓ Mapping the semantics of user with that of content for the purpose of content retrieval, management, creation, etc.

1. Introduction and Background

☐ How Semantic Computing works

- Some areas of semantic computing have appeared as isolated pieces in various fields. Semantic computing glues these pieces together into an integrated theme and addresses their synergetic interactions.



1. Introduction and Background

☐ Current situation of Semantic Computing

- Semantics-based analysis such as data and web mining, analysis of social networks and semantic system design and synthesis have begun to draw more attention from researchers.
- Applications using semantics such as search engines and question answering, content-based multimedia retrieval and editing, natural language interfaces based on semantics have also been attracting attentions.
- Even semantic computing has been applied to areas like music description, medicine and biology and GIS systems and architecture.

Semantic Computing

Making computers and the web more intelligent and interoperable!

1. Introduction and Background

☐ Big Data and Semantic Computing

- The term Big Data has recently gained widespread attention in the field of Information Technology.
- One of the key challenges in making use of Big Data lies in finding ways to uncover relevant and valuable information.
- Semantic technologies make it possible to describe both the data itself and the connections between data sets in ways which are beyond the capabilities of any other technology.
- Semantic computing technologies have been proposed as a means of dealing with these issues, and with the advent of linked data in recent years, have become central to mainstream semantic computing.

1. Introduction and Background

☐ How can semantic computing deal with Big Data(1/3)

- In Big Data, a key problem emphasized everywhere is finding usable data sets, whether in the organization or outside it. Here, semantic technologies can help, since the biggest problem with data sets is often finding out what's in them.
- Semantic technologies make it possible to describe both the data itself and the connections between data sets in ways which are beyond the capabilities of any other technology.

➢ For example IBM's Watson, the data sets produced as part of Linked Open Data can usefully be deployed in deep learning and machine learning contexts.

1. Introduction and Background

□ How can semantic computing deal with Big Data(2/3)

- One of the biggest obstacles for Big Data analytics development is the variety of web-based information, some structured, some numeric, some totally unstructured (e.g. video and audio).
- This is where Semantic Web steps in. It can promote common data formats to make data more consistent and easy to interpret. For Big Data analytics companies, the Semantic Web will allow them to help businesses make even better decisions in real time.

➢ Such as Google starting building this massive knowledge vault over a decade ago and recent years we have seen Google actually start to flex its semantic muscles.

1. Introduction and Background

□ How can semantic computing deal with Big Data(3/3)

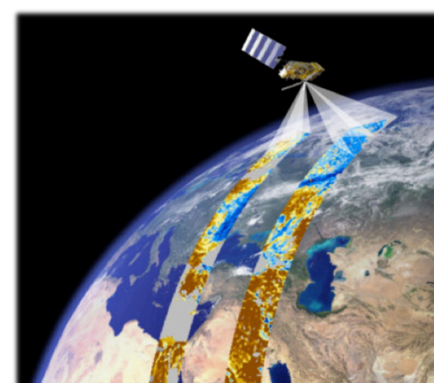
- Semantic technology also makes it much easier to integrate available information for analysis. Suffice it to say that while semantic technologies have a lot to offer in this space, here they are in competition with a more pure Big Data approach.
- For instance TopQuadrant's TopBraid solutions empower people by providing them with connected, business-relevant, actionable knowledge, which transforms diverse information into semantic assets and creates a semantic ecosystem.

TopBraid
Enterprise Vocabulary Net™

1. Introduction and Background

□ **Remote Sensing Data is super abundant**

- From the launch of the first Earth observing satellite, to today's growing space industry, the volume of remote sensing data continues to grow at a remarkable rate. We are on the cusp of a momentous leap forward in data accessibility.
- There are many questions associated with all of this growth: Where will all this data be stored? How will data be efficiently discovered, accessed and visualized? What types of processing and data management tools will be needed? How will this data be used? What new types of applications will be devised to leverage the information derived from this data?



1. Introduction and Background

☐ Remote Sensing Big Data appears

- Big Earth observing data can be defined in terms of (ESA,2013)

➢ volumes

➢ degree of diversity and complexity - including streaming of data from presently available and upcoming satellite capabilities, and innovative ground devices

➢ the unpredictable value added derivable from their innovative analyses and fusion

1. Introduction and Background

☐ Particularity in Remote Sensing Big Data

- From the term big data is relatively new in the field of applied remote sensing and remote sensing data sets are bit different from typical big data sets. Remote sensing time series data sets have both temporal and spatial dimension with multiple ground targets that have different signature in remote sensing signals.
- From the perspective of information processing, remote sensing big data has two types of features: external features (include massive, heterogeneous and multi-source) and internal features (include high-dimensional, multi-scale, and non-stationary).

1. Introduction and Background

☐ Progress in Remote Sensing Big Data

- Early in 2008, IBM proposed the idea of Smarter Planet based on global observation and spatial data.
- ESA gave an event report "Big Data from Space" in June 2013 and setup special committee for the event.
- NASA Brought Earth Science 'Big Data' to the Cloud with Amazon Web Services in November 2013

1. Introduction and Background

☐ Challenges in Remote Sensing Big Data (1/3)

- This huge inflow of remote sensing data will help to develop more efficient monitoring and decision support systems with high consistency. Then new frame works are required to handle this new upcoming inflow of remote sensing data with fine resolution and wide coverage. This frame should support:

➢ A common ground with respect to regional and global applications and the integration of different data sources

➢ Data intensive and innovative services across diversity and heterogeneous resources.

➢ A baseline of activity to make the identified scenarios actionable and practicable.

1. Introduction and Background

☐ Challenges in Remote Sensing Big Data (2/3)

- The difficulty lies in the data storing, memory loading, processing and also analyzing.

➢ Efficient representation and modeling for geospatial big data

➢ Efficient storage and management of the massive RS data

➢ The intensive irregular data access patterns charge for the poor parallel I/O performance

➢ The loading and transmission of RS big data

➢ Tons of data-dependent tasks for optimal scheduling

➢ Efficient and productive programming of RS applications on hierarchical cluster-based parallel system.

1. Introduction and Background

☐ Challenges in Remote Sensing Big Data (3/3)

- The development of new spatial indexing and algorithms to handle real-time, streaming data and to support topology for real-time analytics.
- The development of conceptual and methodological approaches to move big data from descriptive and correlation research to ones that explore casual and explanatory relationships.
- The development of efficient methods to display data integrated in the three dimensions of geographic and one dimension of continuous time.
- The development of novel approaches for error propagation so as to effectively assess data quality requires.

1. Introduction and Background

☐ Semantic Computing in RS Big Data

- Among these challenges above, one important and crucial is how to know what and how data to use, particularly for people new to remote sensing.
- There is free data, paid for data, various satellites, various data types, various formats and the list can go on. The remote sensing community needs to help by providing more bridges between the data and the user community.
- The datasets available can offer huge benefits for business and science, but if people have to spend hours hunting round and trying to find the right image for them, they won't stay users for long.

1. Introduction and Background

☐ Roles of Semantics in remote sensing application

- Semantics don't specify how to use, where to use and who use it, but help to know what kind of data based on its inherent characteristic, which will help to judging how to use, and where to use and who use it in various environments. It can make big data smarter.

✓ Provides end-users increased ability to self-manage data from varied sources

✓ Addresses varying user needs and changing in different application environments

✓ Manages terminology, concepts and relationships while connecting diverse data from varied data sources

1. Introduction and Background

□ Points of Semantic Computing in RS Big Data

- Efficient representation and modeling for geospatial big data (describing both the data itself and the connections between data sets in way of linked data based on domain knowledge)
- Analyzing, mining and visualizing geospatial big data for decision-support
  (supporting semantic inference and associative retrieval to find latent link and potential value out of mass data)
- Quality assessment of geospatial big data from new sources
  (checking for inconsistencies and uncertainties based on currently built knowledge base and known domain rules)

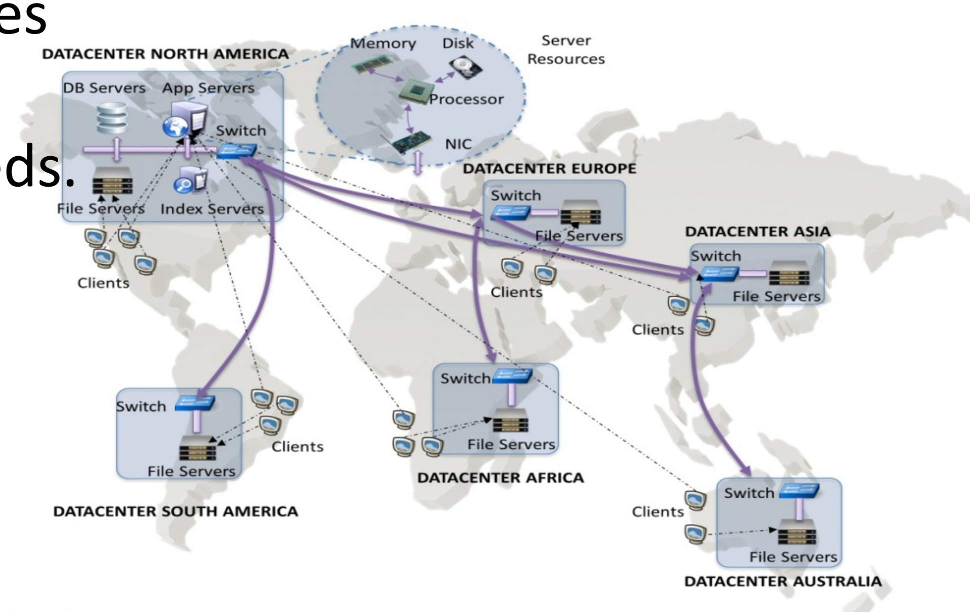
1. Introduction and Background

■ **Overview of semantic computing in Big Data challenge**

- The trend of larger data sets is due to the additional information derivable from analysis of a single large set of related data, compared with several smaller sets with the same total amount of data, thus allowing correlations to be found.
- Semantics don't specify how to use, where to use and who use it, but help to know what kind of data based on its inherent characteristic, which will help to judging how to use, and where to use and who use it in various environments.
- The challenge of Big Data stimulates the developm
semantic computing, and also Big Data requires th

2. Remote Sensing Big Data Semantic Organization

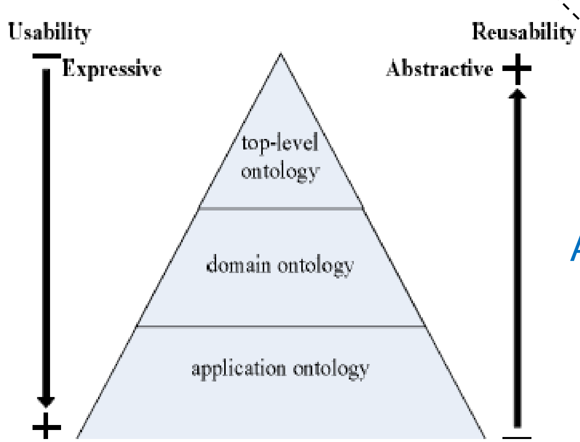
- Different data sources can define the same entity, concept or term differently. There is a need not only to have a glossary of terms and entities but also to manage the relationships between different data and meta-data so that search, data lineage and other actions can be performed.
- Semantic technique addresses the data relationships and meta-data management needs.
- A smart data layer can be placed over big data to manage relationships across all these varied sources.

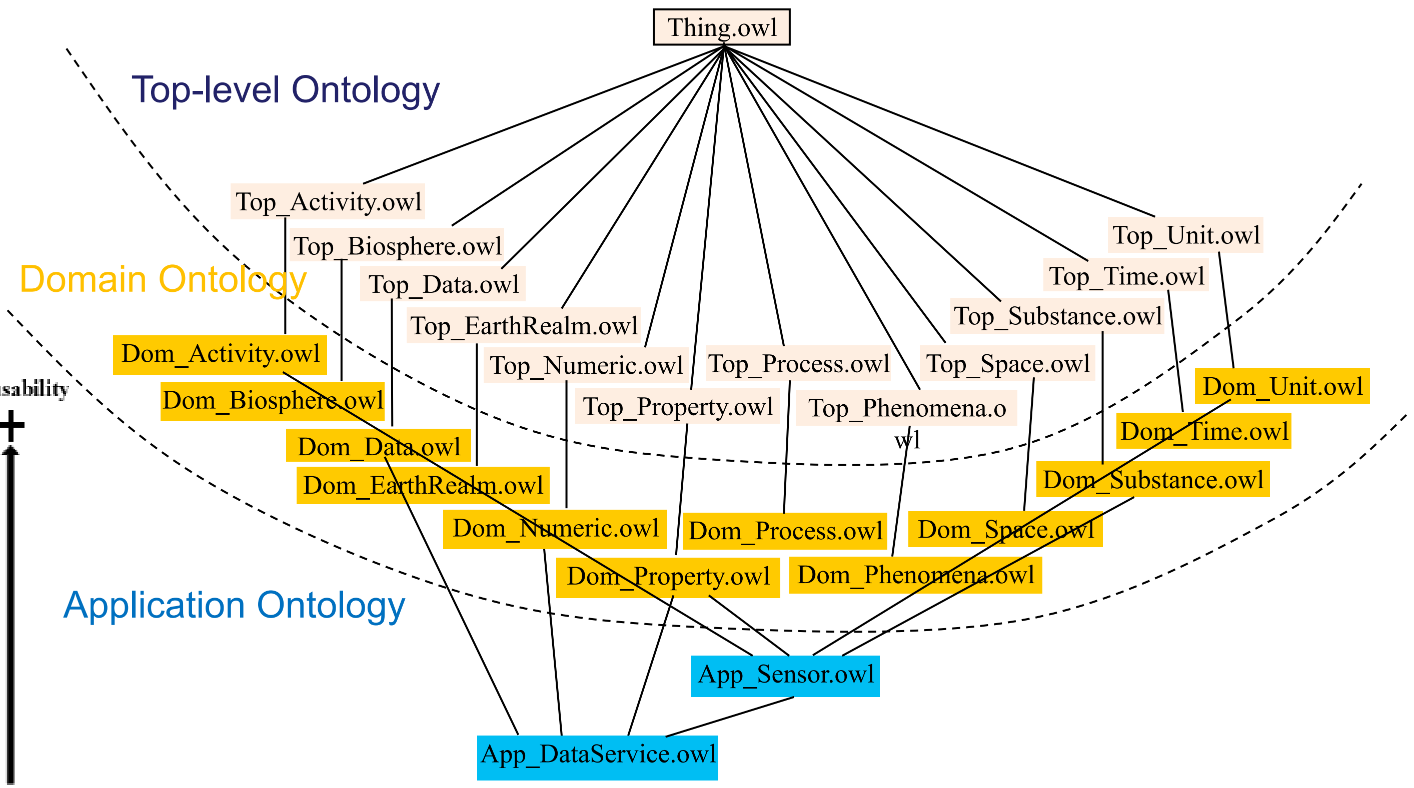


2. Remote Sensing Big Data Semantic Organization

☐ Manage terminology, concepts and relationships among diverse data sources

- Built a three-layer hierarchical spatial information ontology

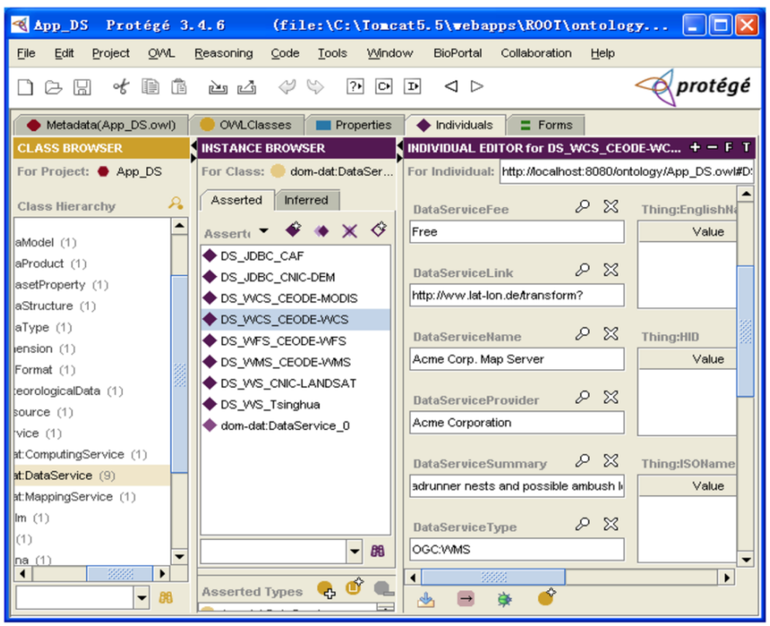


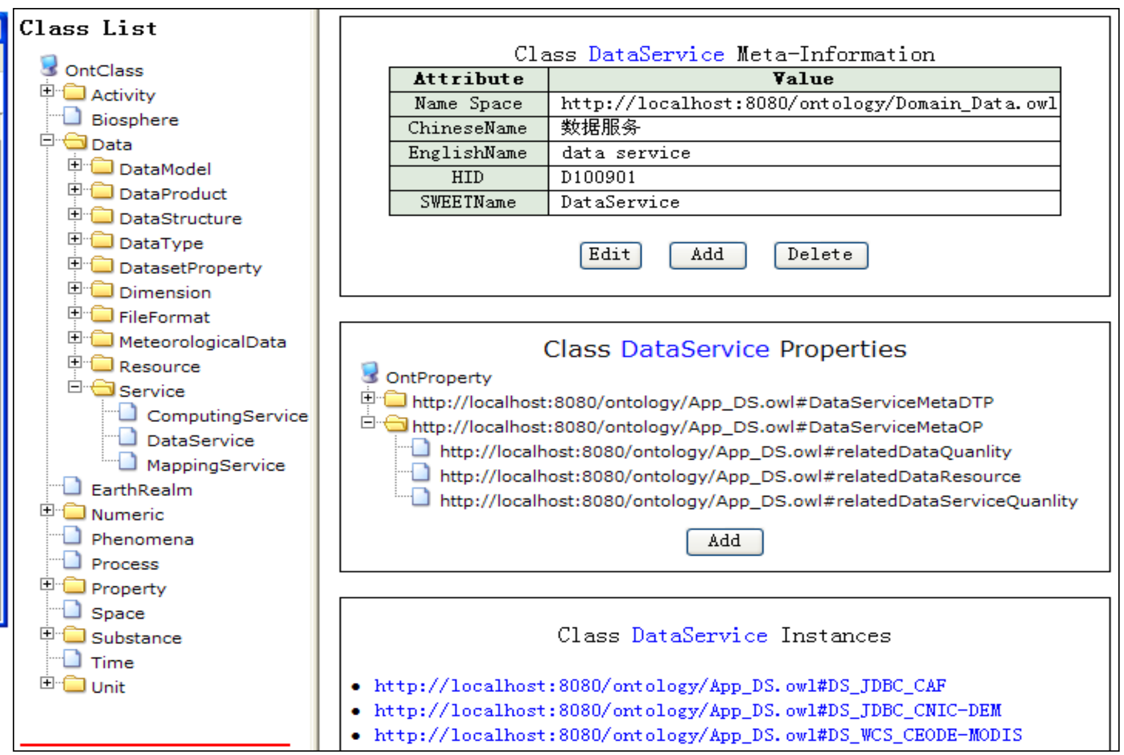


2. Remote Sensing Big Data Semantic Organization

☐ Manage terminology, concepts and relationships among diverse data sources

- Described variant data sources by semantic property

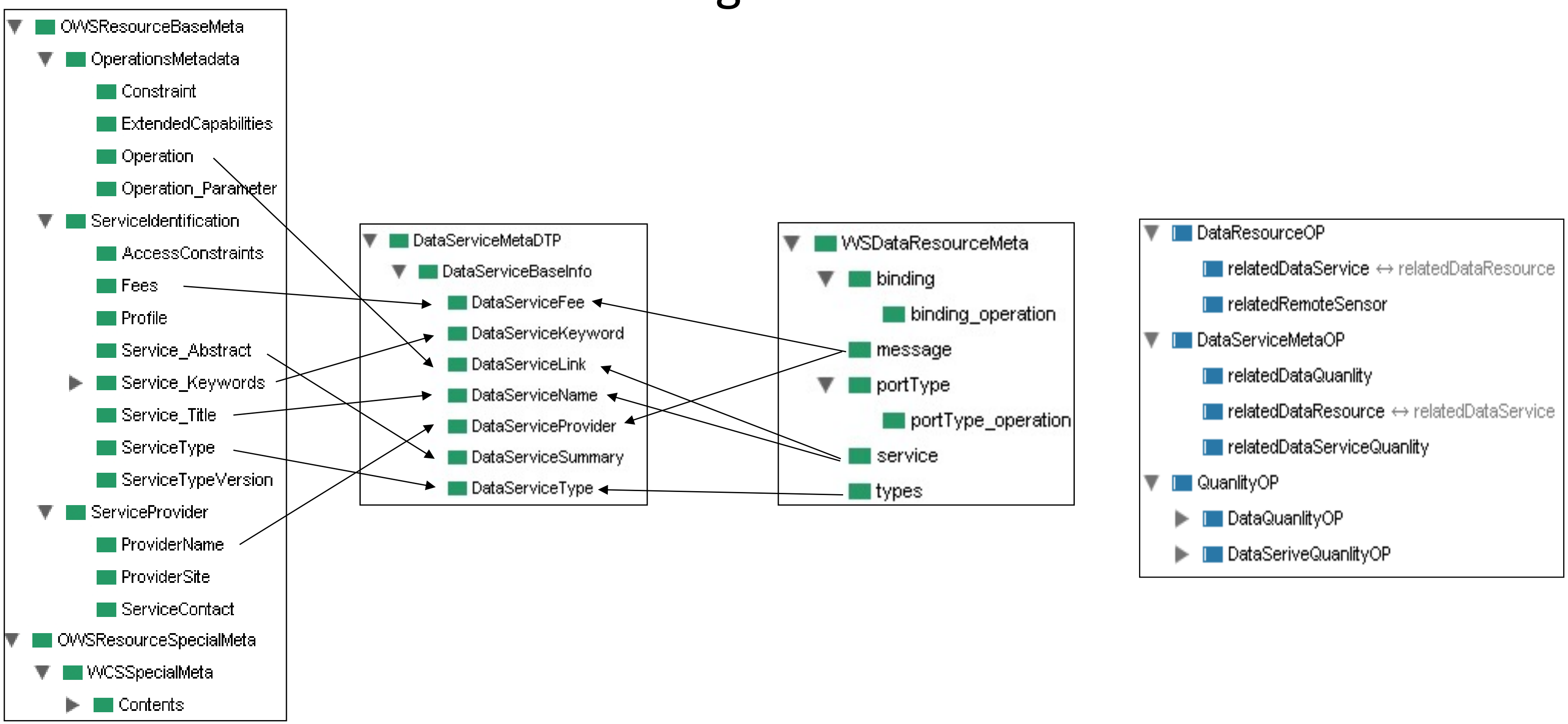




2. Remote Sensing Big Data Semantic Organization

☐ Manage terminology, concepts and relationships among diverse data sources

- Made semantic links among different meta-data of data

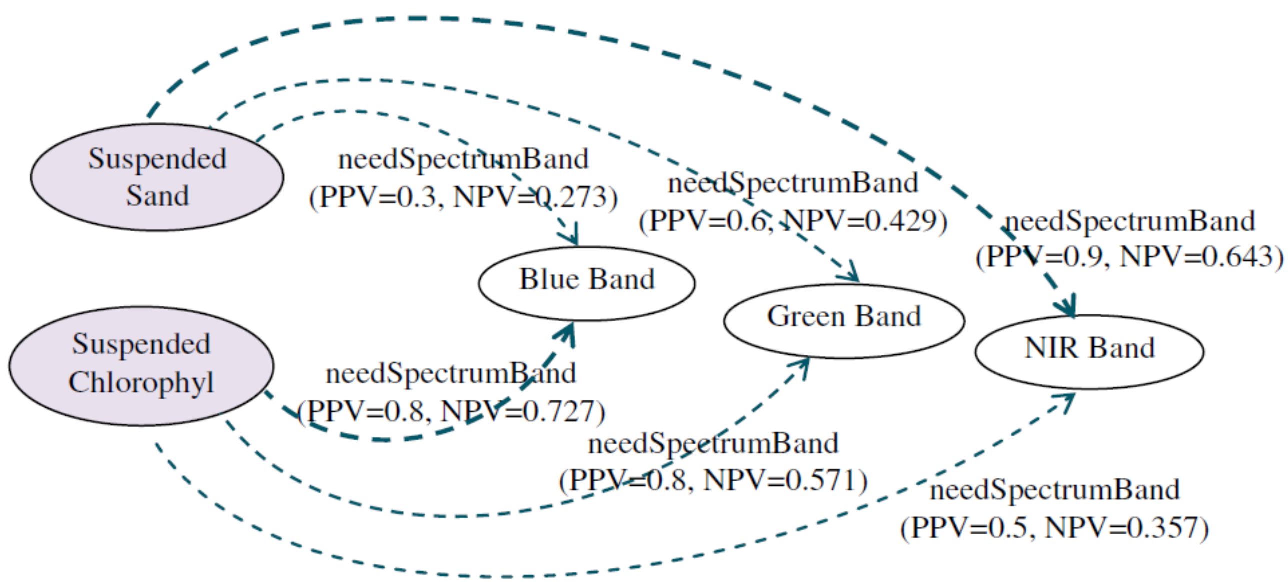


2. Remote Sensing Big Data Semantic Organization

☐ Measure and describe uncertainty in RS big data

- Introduced Possibility and Necessity degrees to measure various uncertainties

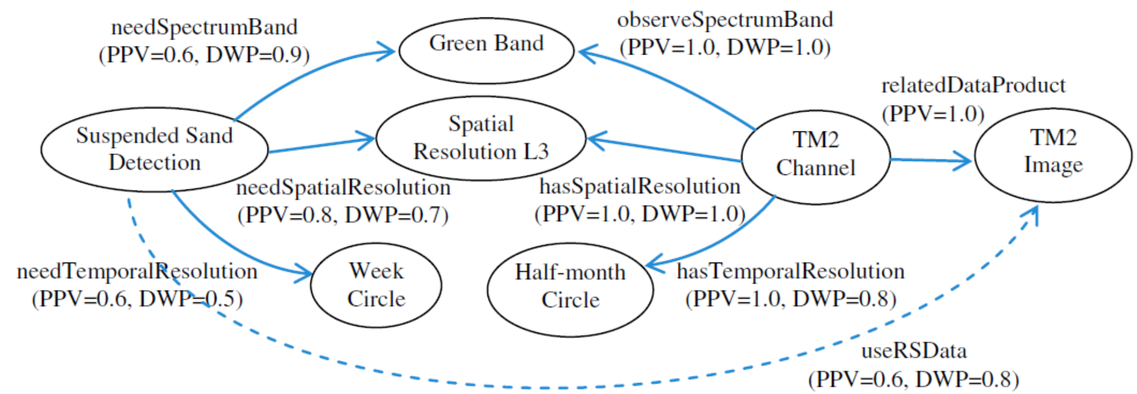
$$\mathrm{Pos}(r) = \max\{\pi_i(r) : \pi_i(r) \in \Pi\} \qquad \mathrm{Nec}(r) = \mathrm{Poss}(r) / (\textstyle\sum \mathrm{Poss}(r') + \mathrm{Poss}(r))$$

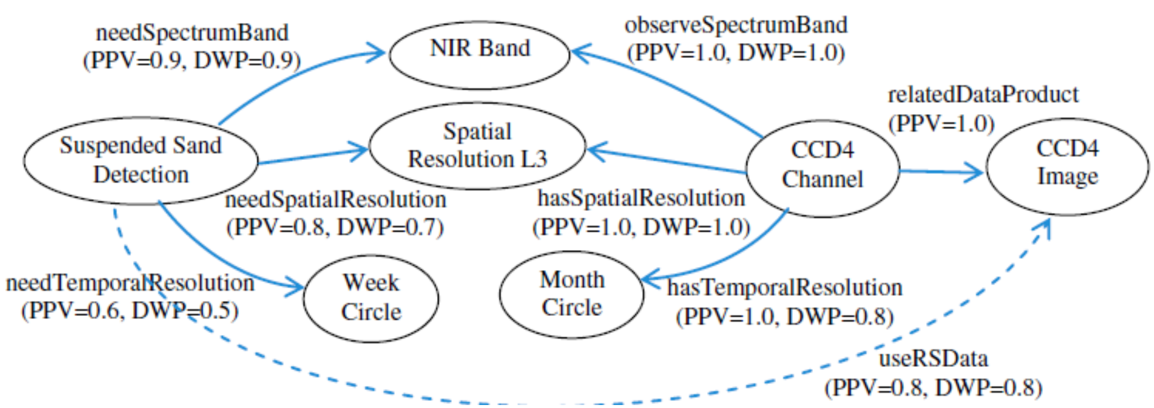


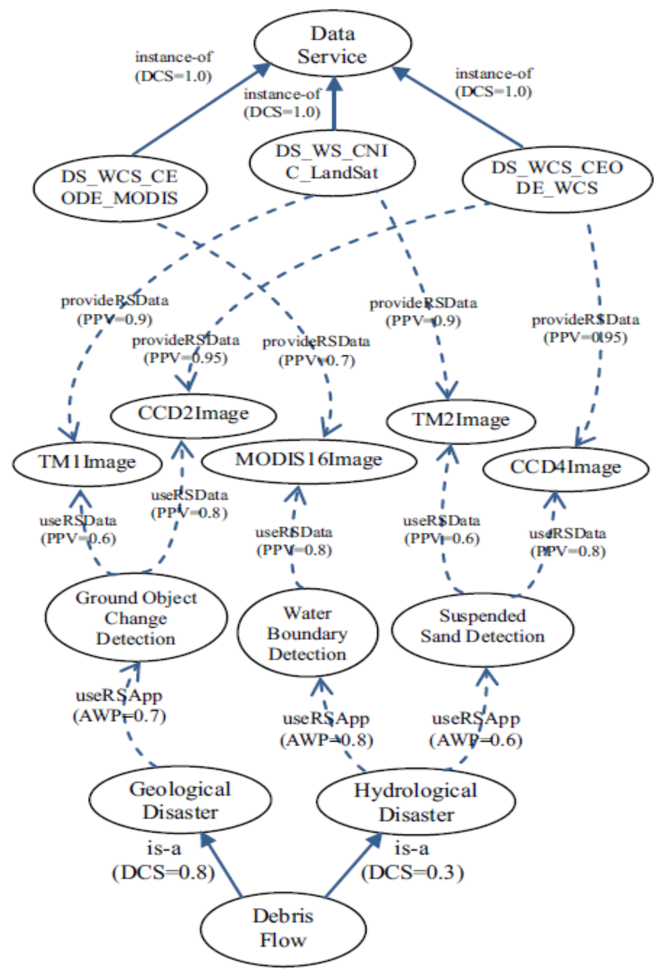
2. Remote Sensing Big Data Semantic Organization

☐ Measure and describe uncertainty in RS big data

- Measured different semantic relations with the quantitative measurements

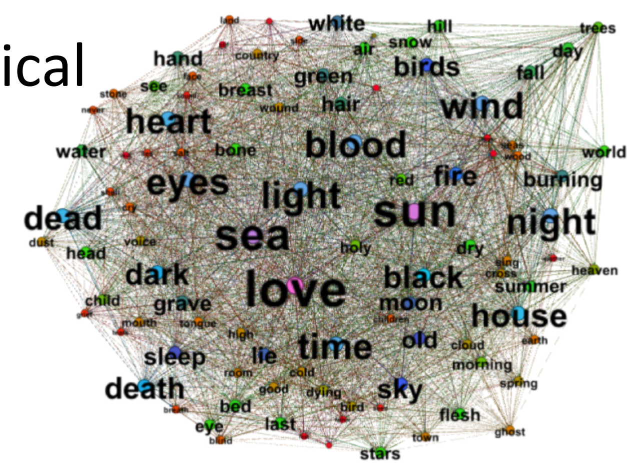






3. Remote Sensing Big Data Information Retrieval

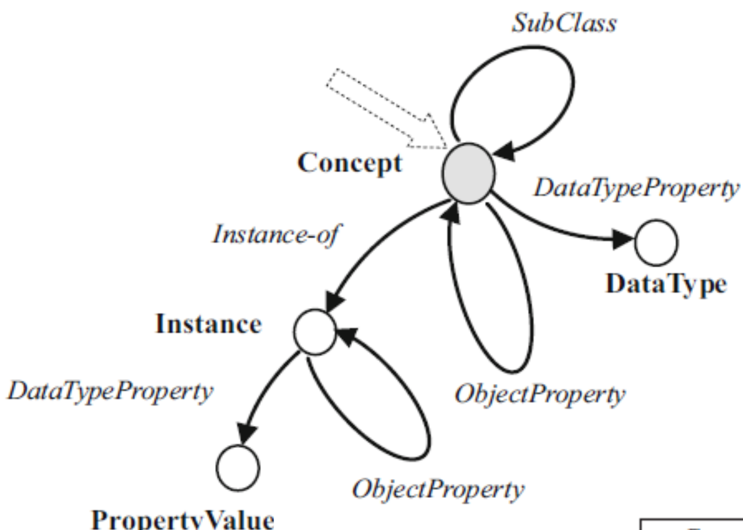
- Traditional processing and application methods for spatial data mostly obey the process of causal analysis based on strict mathematical formulas, which typically endeavor to find all the causes and effects of everything.
- In remote sensing big data, this type of data usage is no longer applicable. Users usually wish to obtain all possible related resources, and care less about why or how these correlations occur or on which classical theories they are based.
- Therefore, some imprecise and non-classical methods are needed to realize a similar fuzzy retrieval of spatial big data based on semantics.

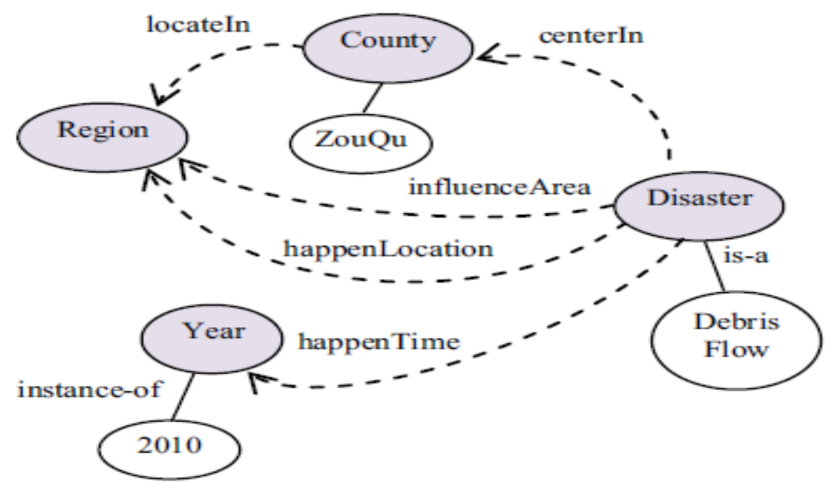


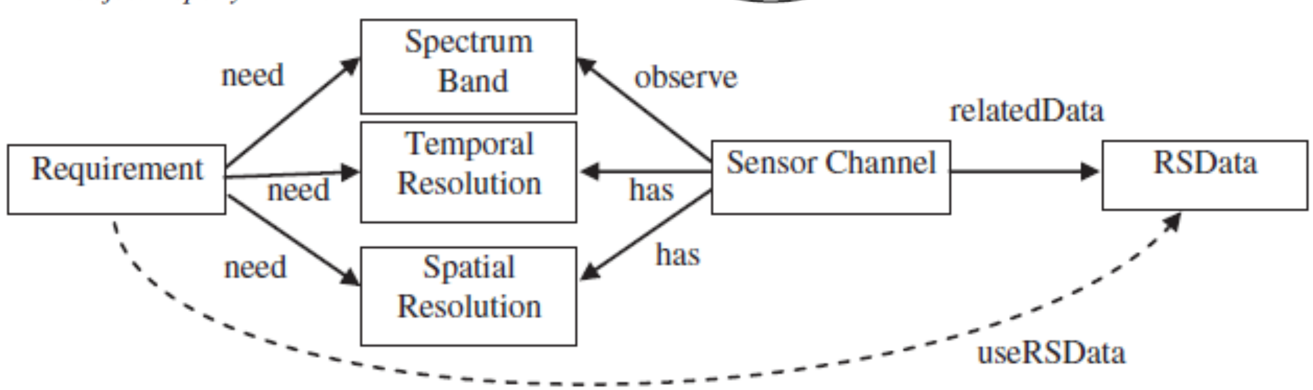
3. Remote Sensing Big Data Information Retrieval

□ Semantic retrieval of spatial data and service

- Relevant links among entities in RS data based on semantic



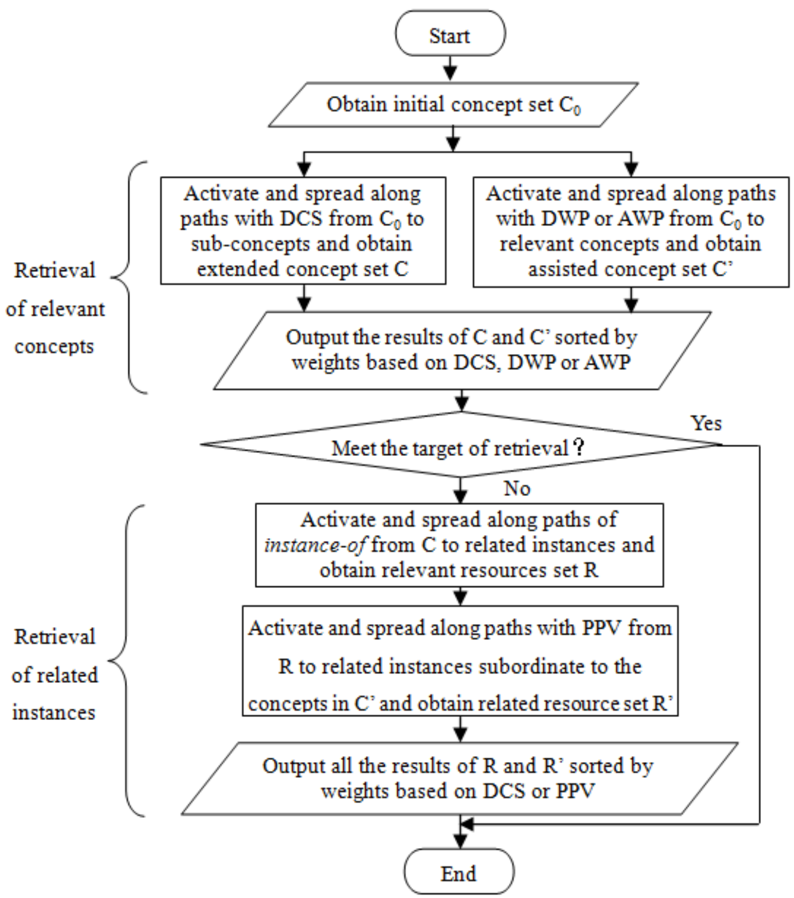


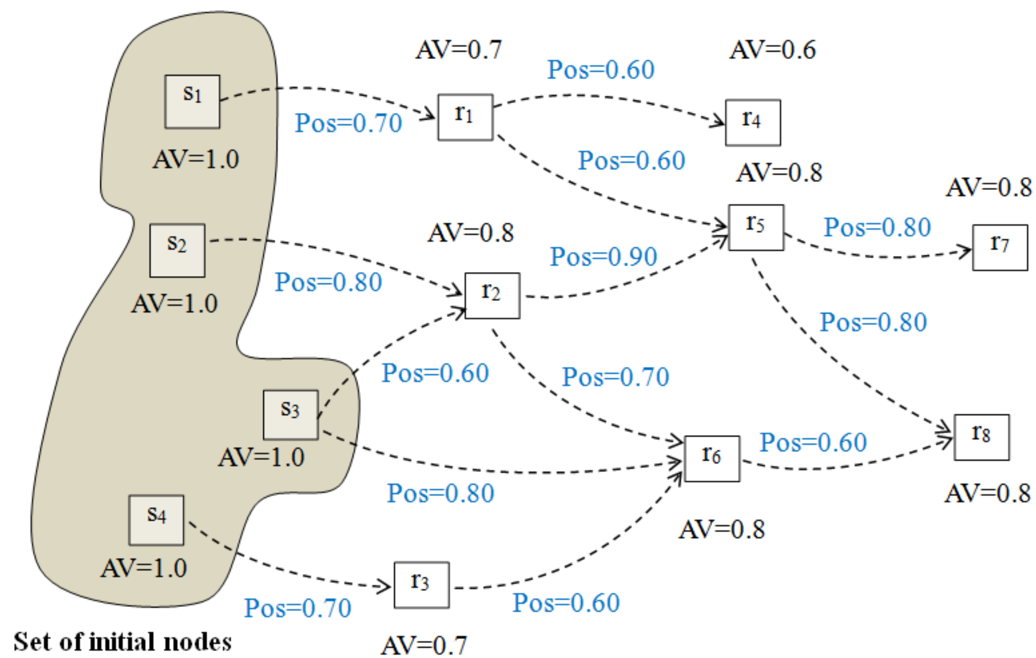


3. Remote Sensing Big Data Information Retrieval

☐ Semantic retrieval of spatial data and service

- Activation - spreading process in semantic ontology space

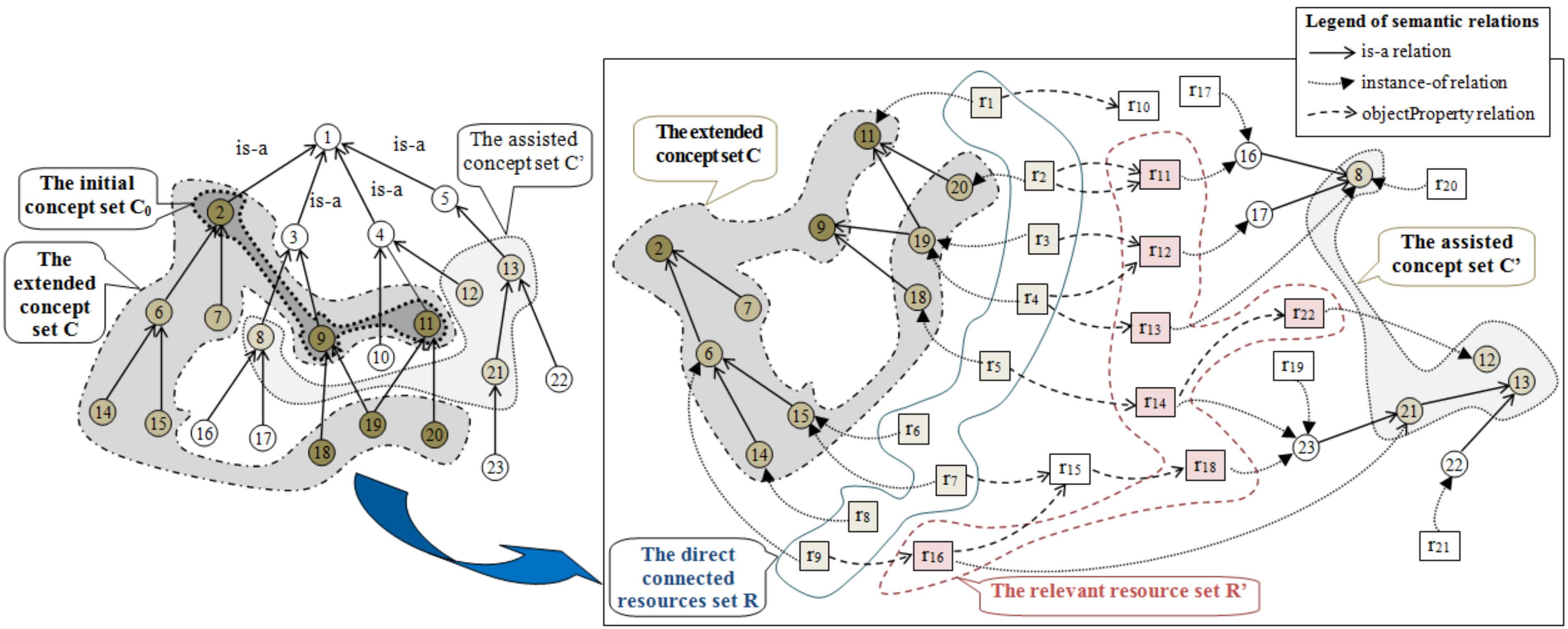




3. Remote Sensing Big Data Information Retrieval

☐ Semantic retrieval of spatial data and service

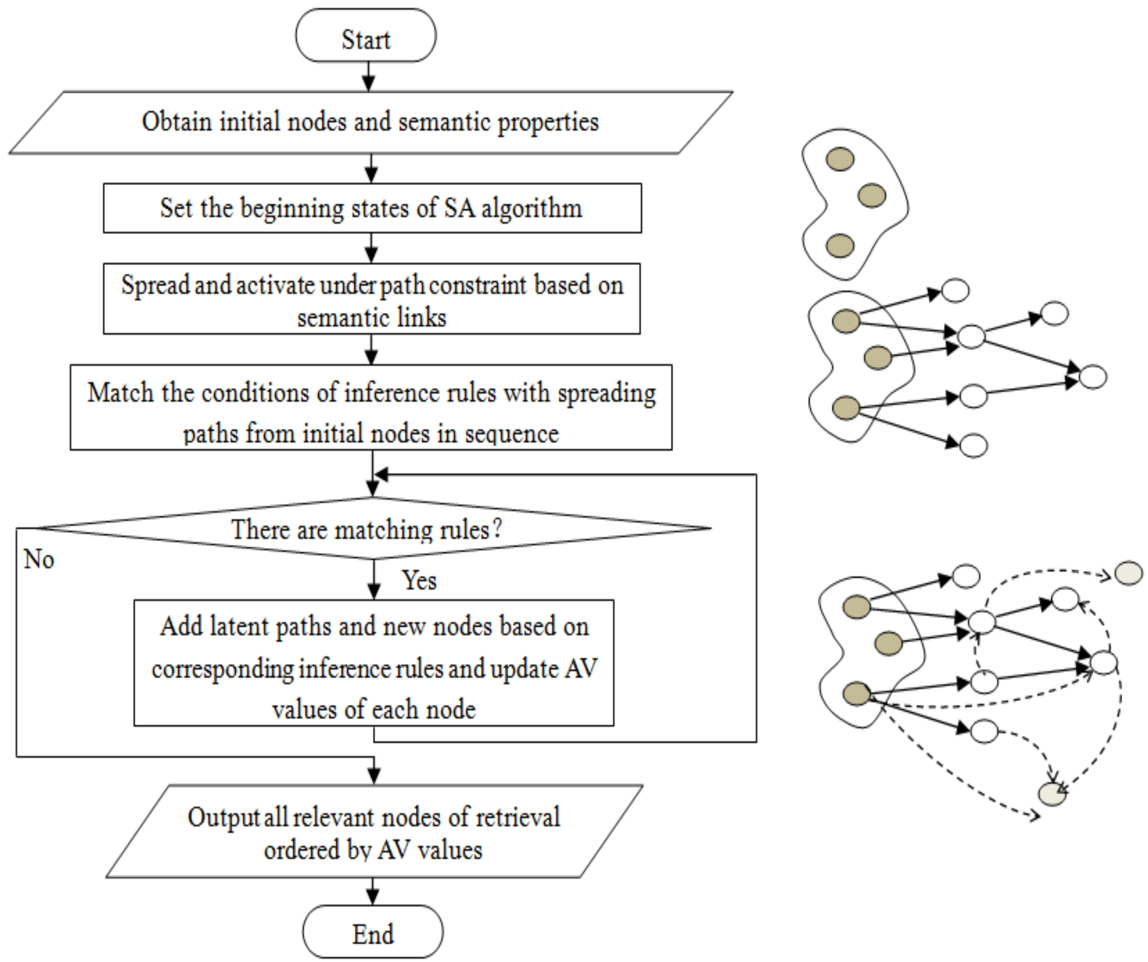
- Associative retrieval based on Spreading Activation algorithm



3. Remote Sensing Big Data Information Retrieval

☐ Combination of Activation Spreading and Semantic Inference

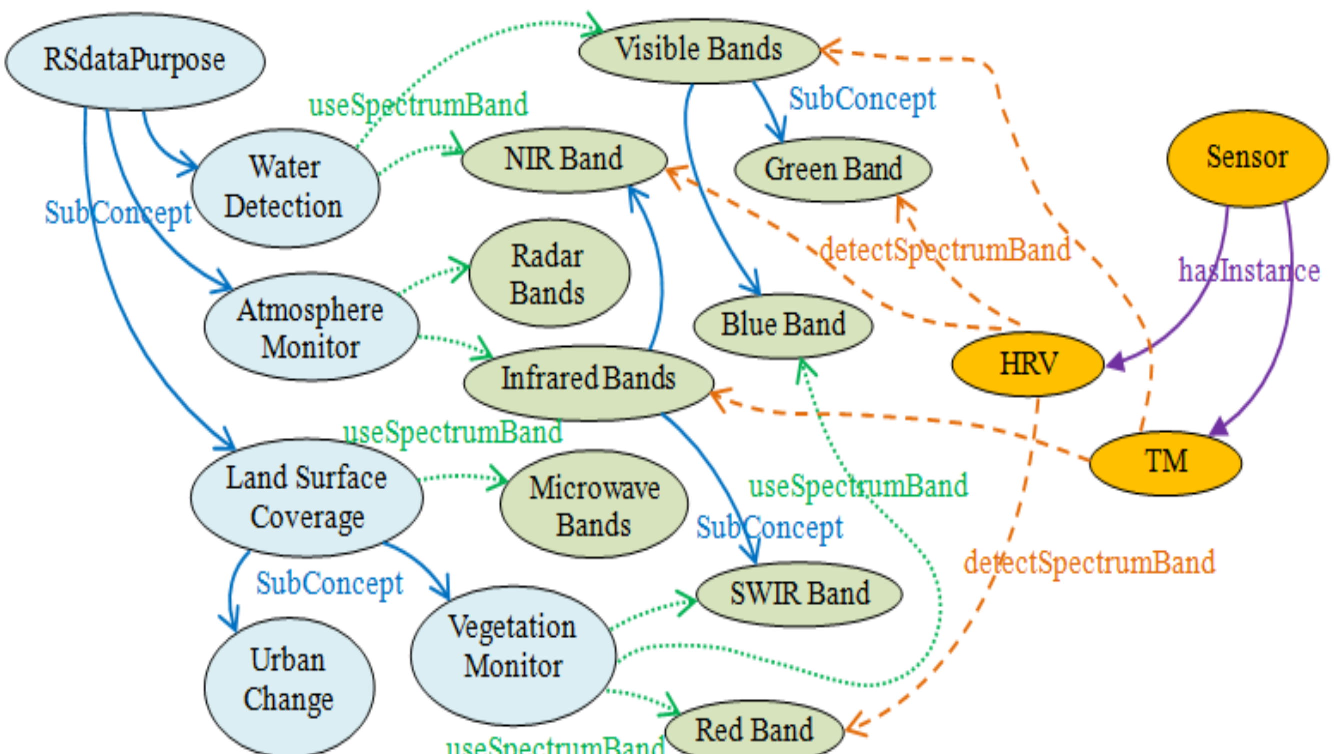
- Find latent paths based on semantic inference rules in SA algorithm



3. Remote Sensing Big Data Information Retrieval

☐ Combination of Activation Spreading and Semantic Inference

- Supplement potential searching paths in associative retrieval

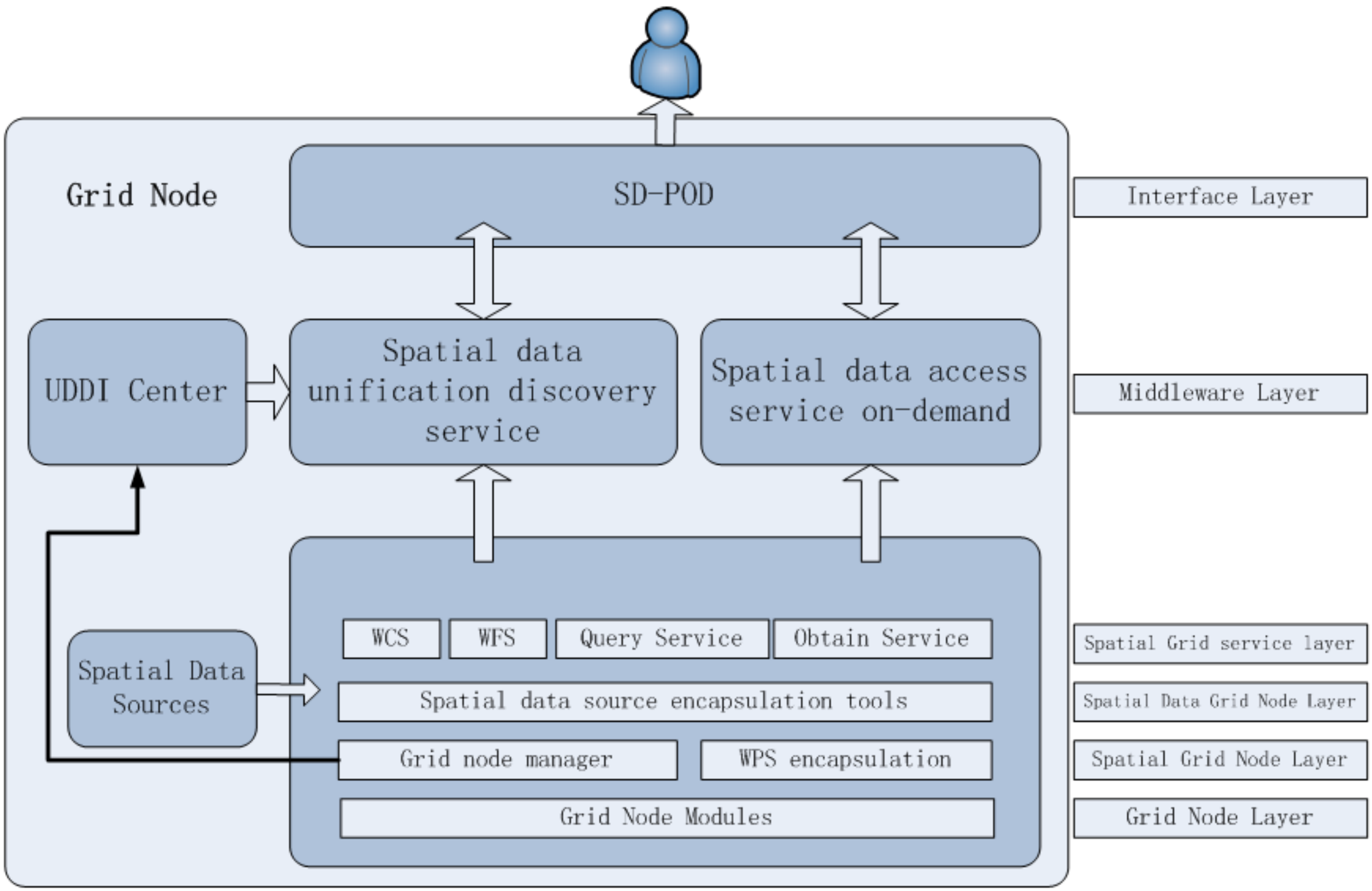


4. Remote Sensing Big Data Semantic Service

- In traditional big data IT solutions, the data model and the IT solutions are designed to address specific business needs and to handle specific data types and data sources.
- Semantic-based solutions have data models that can evolve in run time. This allows the solutions to evolve with user customization requirements and improve business environments.
- SOA infrastructure over existing data stores allows in run time to bring in data into the big data store as necessary. A user can start with something quick and then evolve the solution, adding new datasets as needed, saving significant support time and expenses.

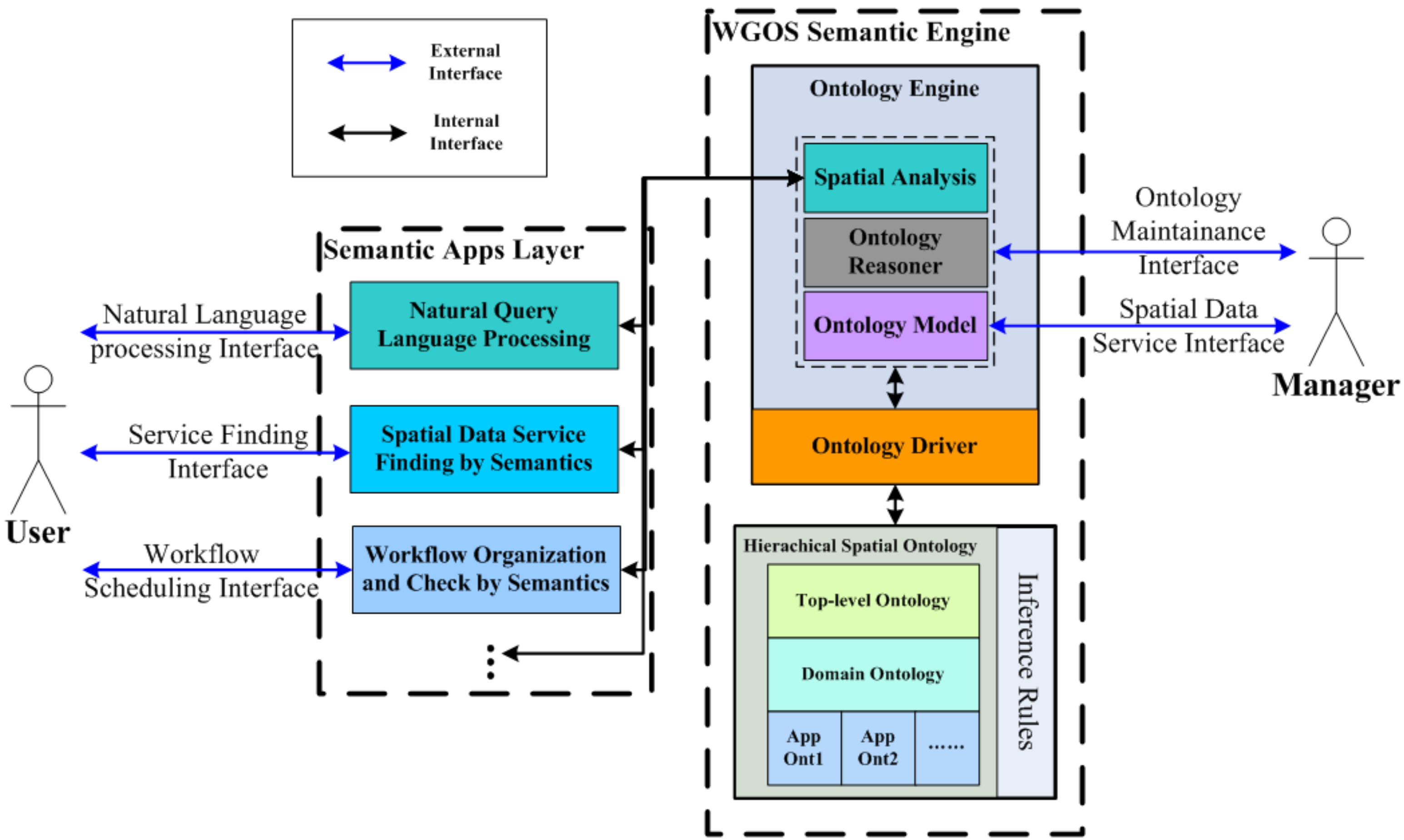
4. Remote Sensing Big Data Semantic Service

☐ Heterogeneous Spatial Data On-demand Service



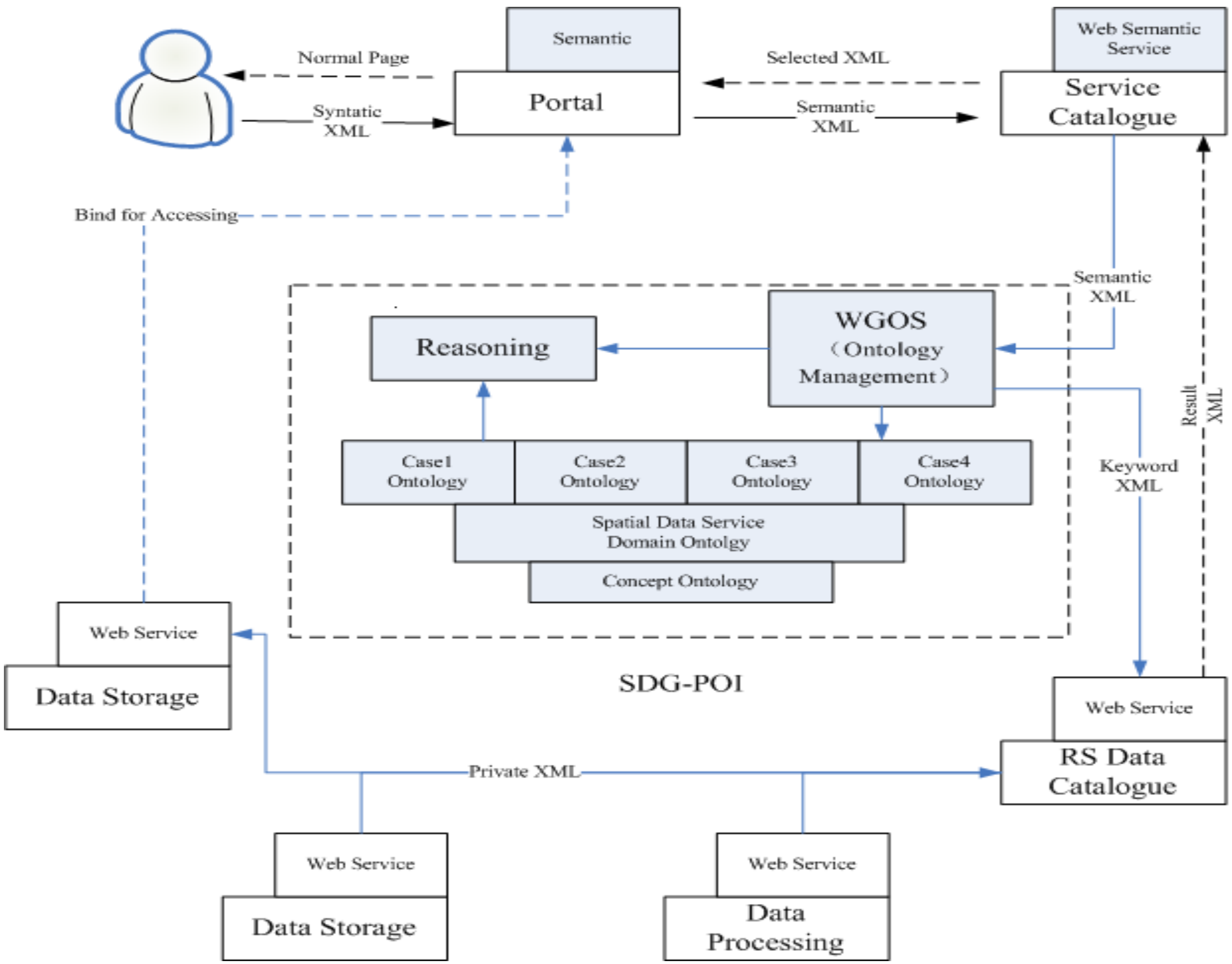
4. Remote Sensing Big Data Semantic Service

☐ Web Geo-Ontology Service Framework



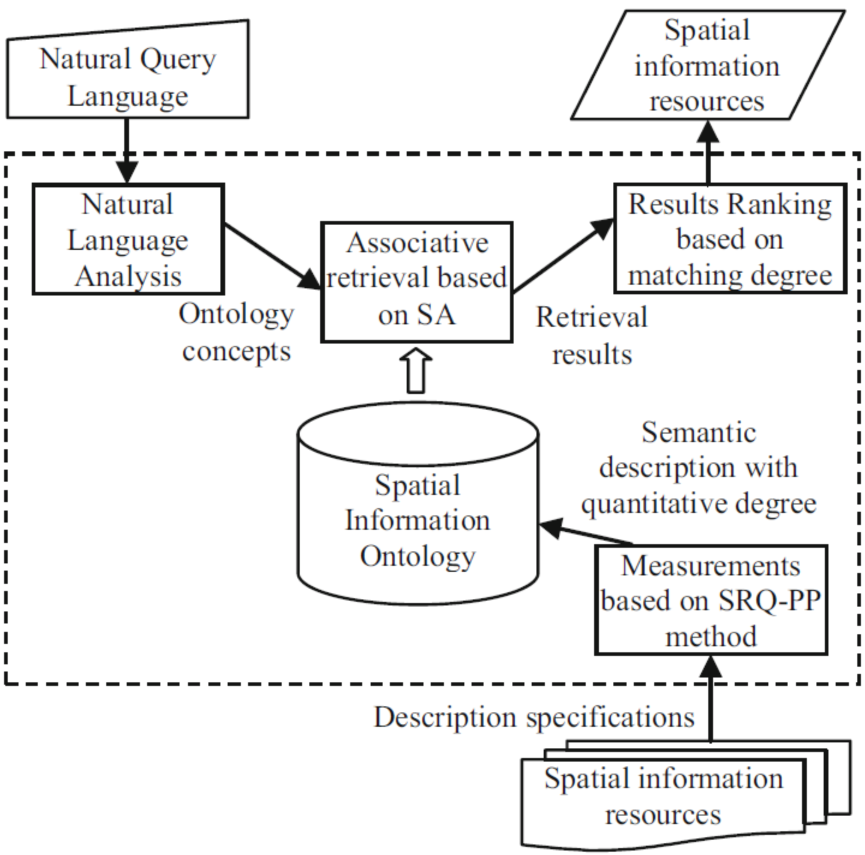
4. Remote Sensing Big Data Semantic Service

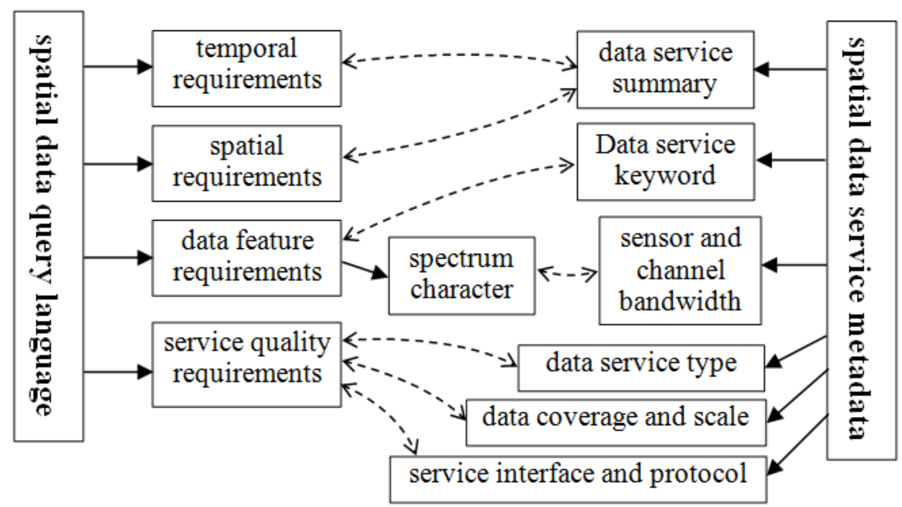
☐ RS data semantic search in Spatial Information Grid



4. Remote Sensing Big Data Semantic Service

☐ Spatial data service query by semantic association





No.	Data Source	Provider	Data Type	Data Size	Weight
1	CEODE-WCS	CEODE	Landsat ETM	2.53 GB	0.8
2	CEODE-MODIS	CEODE	HDF	3600.47 GB	0.7
3	CNIC-Landsat	CNIC	Geotiff	5031.08GB	0.6
4	Tsinghua-MODIS	Tsinghua	HDF	20.30GB	0.3

5. Summary

- The role and value of semantics in Big Data present at the essence positioning and association analysis of variant data, which provides high level abstraction of the overall organization of knowledge across domains and a clear perspective of the value mining in Big Data.
- Semantic computing is not the only or best tool for dealing with Big Data, but we think it may be an efficient investment currently.
- It will make Big Data more smart and convenient.

Open Questions

- Data quality
  - The biggest weakness for semantic technologies in this regard may be that most of the approaches in use so far assume clean data.
  - Reasoner, query engine and so on do reasoning without taking into account that data may be wrong, contradictory, and uncertain.
  - Machine learning techniques have done much better at this so far. Then how to join them with semantic computin efficiently and practically.

Open Questions

- Make use of RS data “semantic enabled”
  - Associative retrieval and linked data has play roles in tackling Big Data.
  - But how to connect, link and make available all data on the Web that is created in large volume at high-speed in different variety and variability ensuring the correctness and quality of that data and making it understandable for humans and machines.
  - With the increasing of semantic information, Semantic Web will also become a kind of semantic big data.
  - How to label the RS data unlabeled

Open Questions

- Develop “adaptive” and “smart” models
  - The existing ontology knowledge framework is fixed and static, which need to adapt to the dynamic changes of structure adjustment automatically.
  - Current ontology model is some complicated and not easy to use, to support cross domain knowledge sharing and incremental updating requirement ontology model to have more simple but extendable and accessible structure.
  - Ontology model need further evolution and maintenance mechanisms in wide distributed environment.

Further reading

- Shengtao Sun, Lizhe Wang, Rajiv Ranjan, Aizhi Wu: Semantic analysis and retrieval of spatial data based on the uncertain ontology model in Digital Earth. Int. J. Digital Earth 8(1): 1-14 (2015)
- Shengtao Sun, Lizhe Wang, et. al. Association Retrieval in Spatial Big Data based on Spreading Activation and Semantic Ontology. Submitted to Future Generation Computer Systems.