

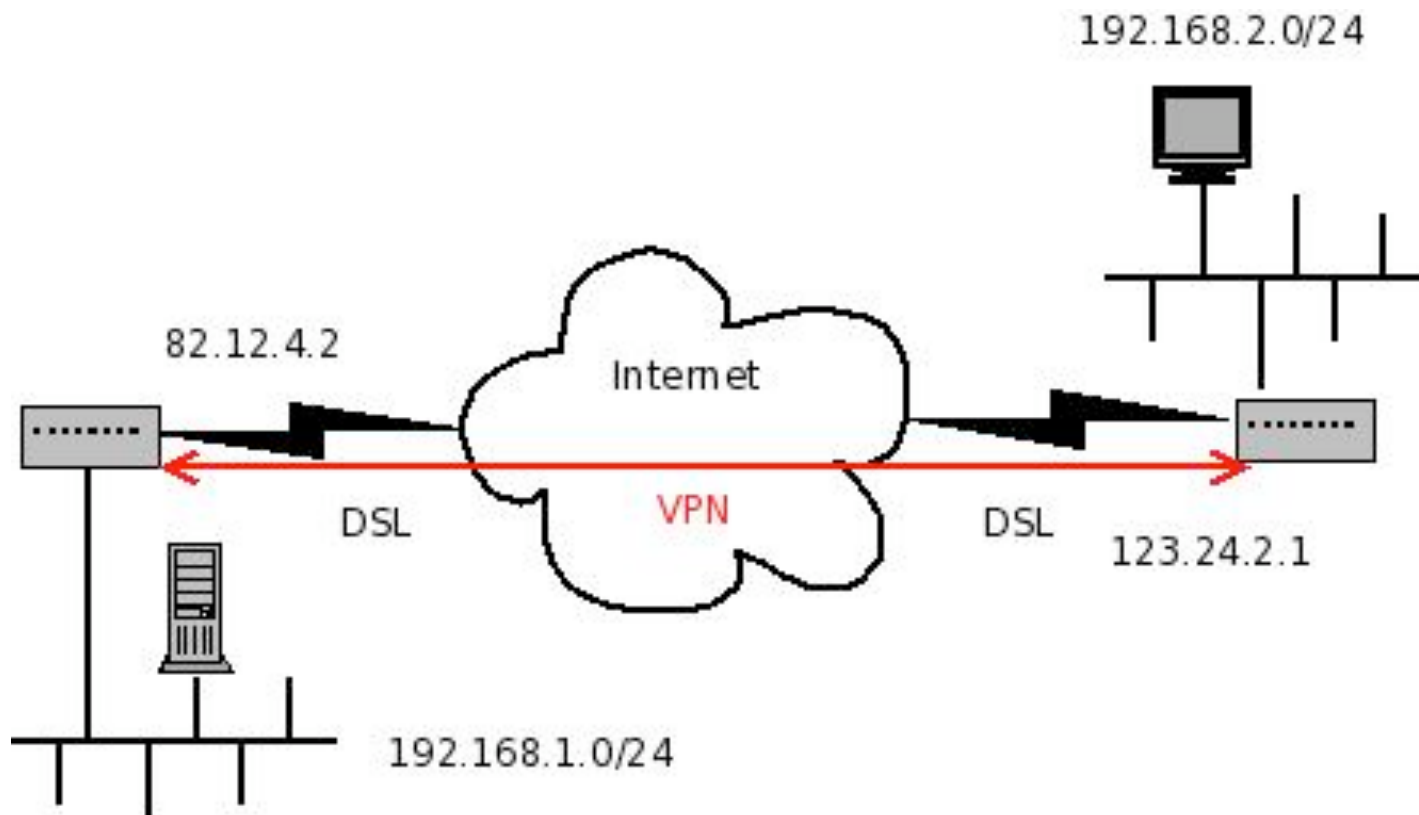
Semantic Interpretation of Internet Monitoring Data

Ian Napier

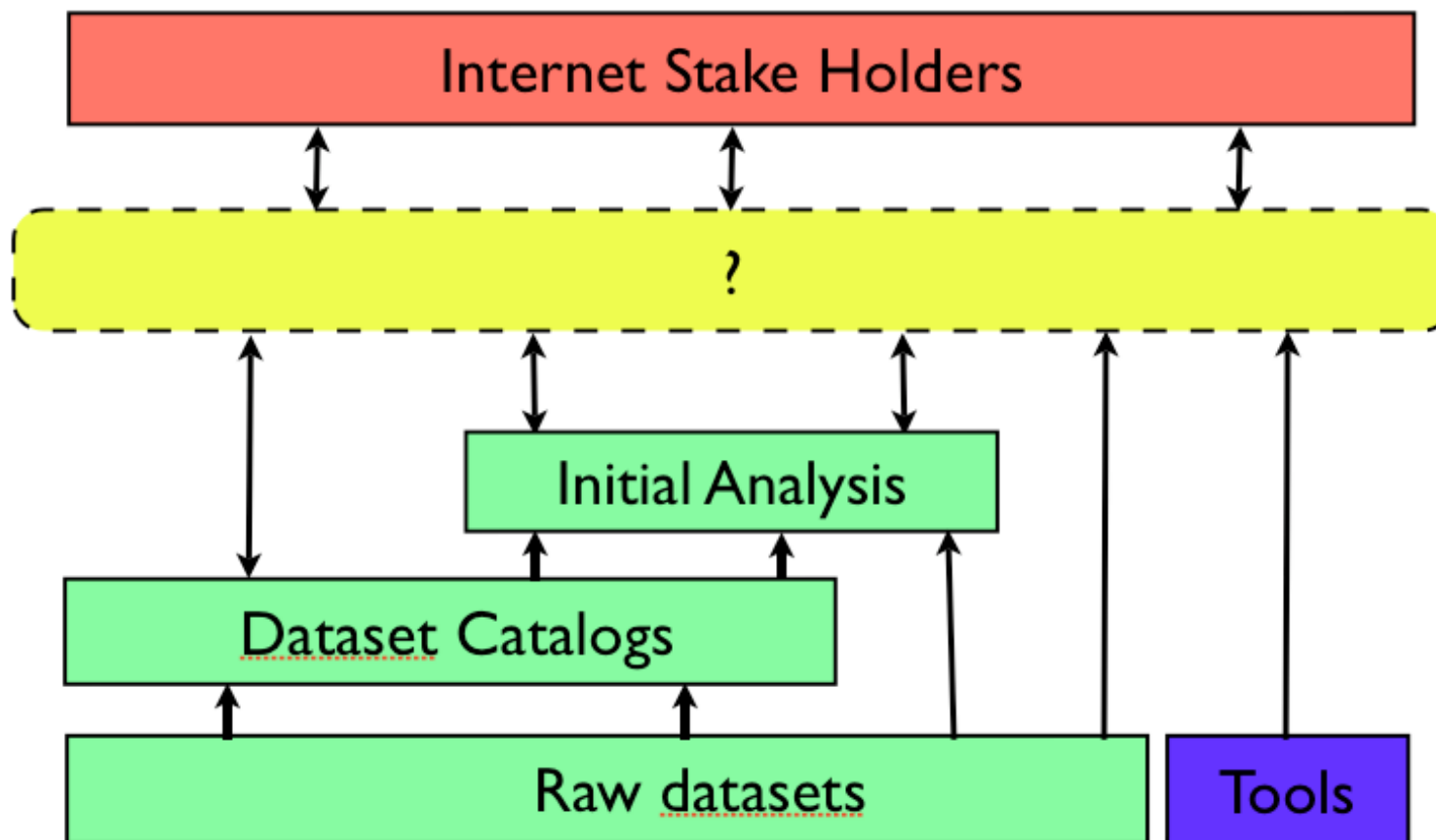
PhD Student, Dept of Computer Science
Loughborough University

Supervisors : Dr Iain Phillips
Dr Lin Guan

Example Scenario : Intersite VPN



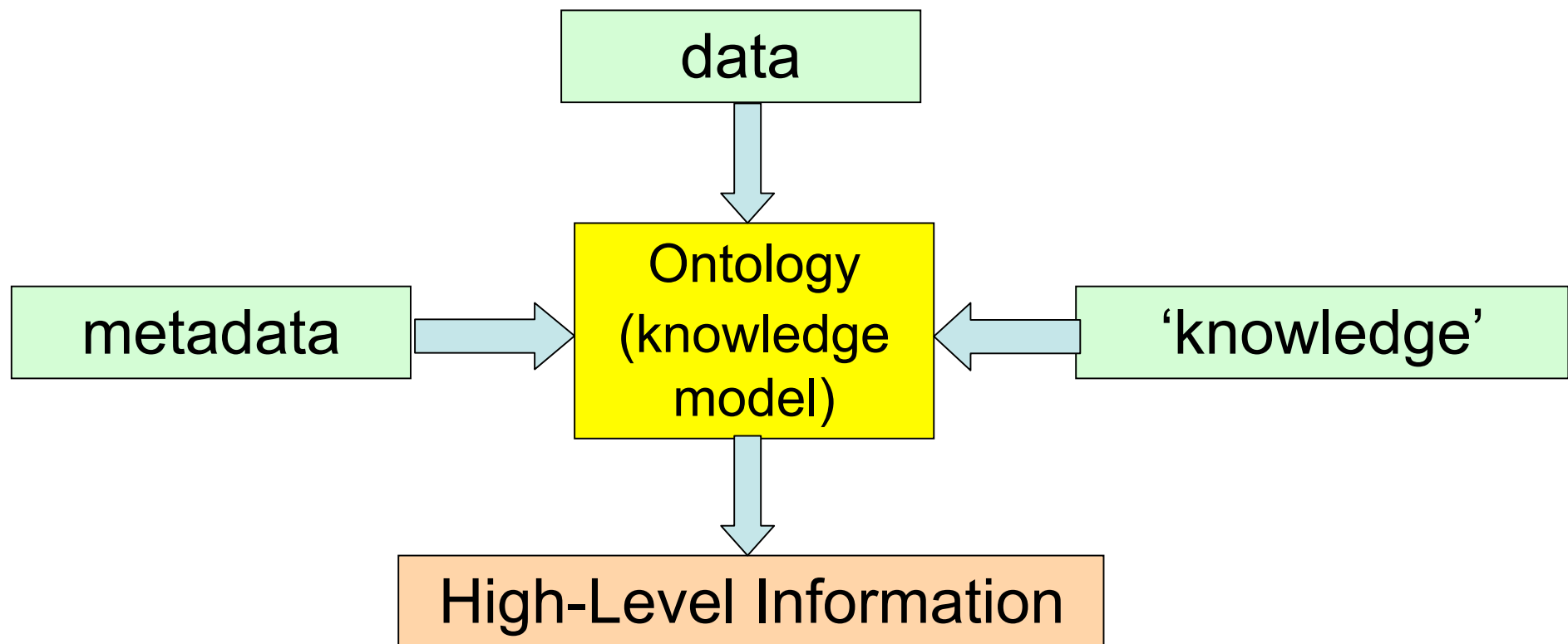
Where's the gap?



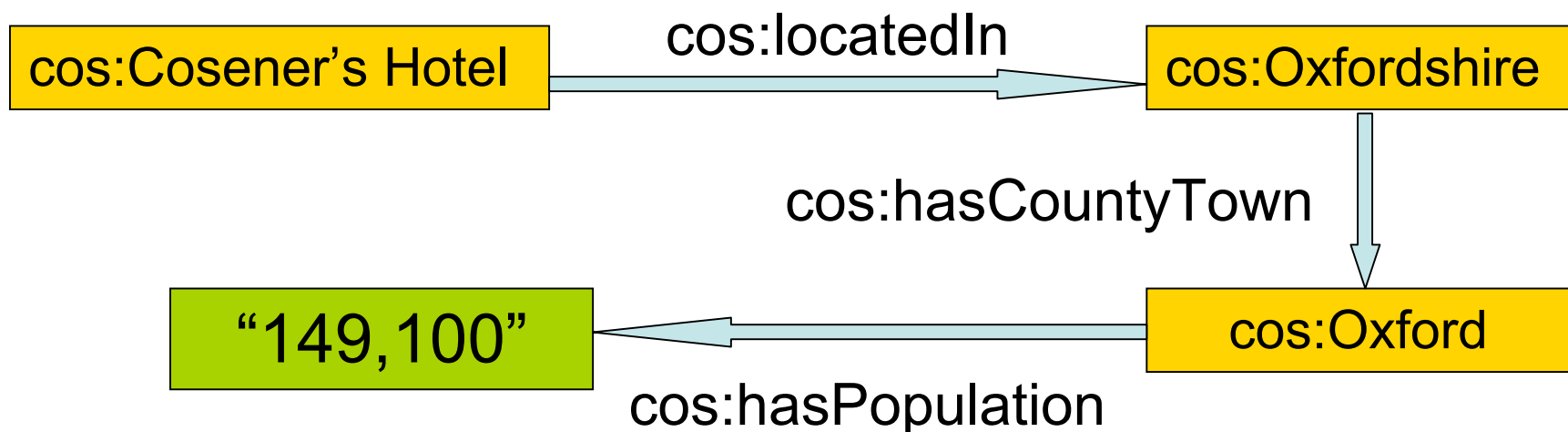
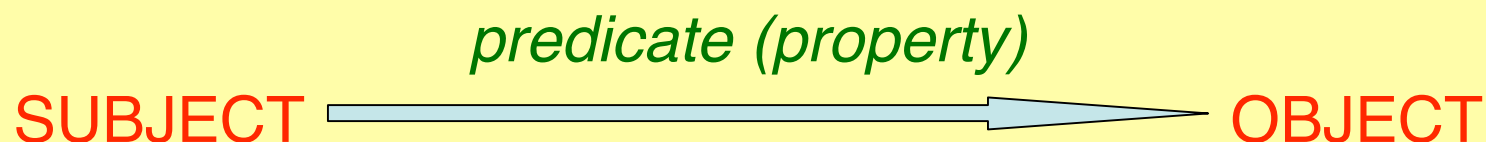
- Different published formats (PCAP files, CSV, HTML, spreadsheets, databases etc.) **How to access and combine these?**
- Inconsistent or Conflicting data (eg : AS level topology from BGP updates or Active probing methods [4]). **How to mediate?**
- Understanding the significance of, and relationship between disparate data. Interpretation and semantics need domain expertise. **How to encapsulate this knowledge?**
- **RDF data models and Ontologies are one approach.**

A nice general definition :

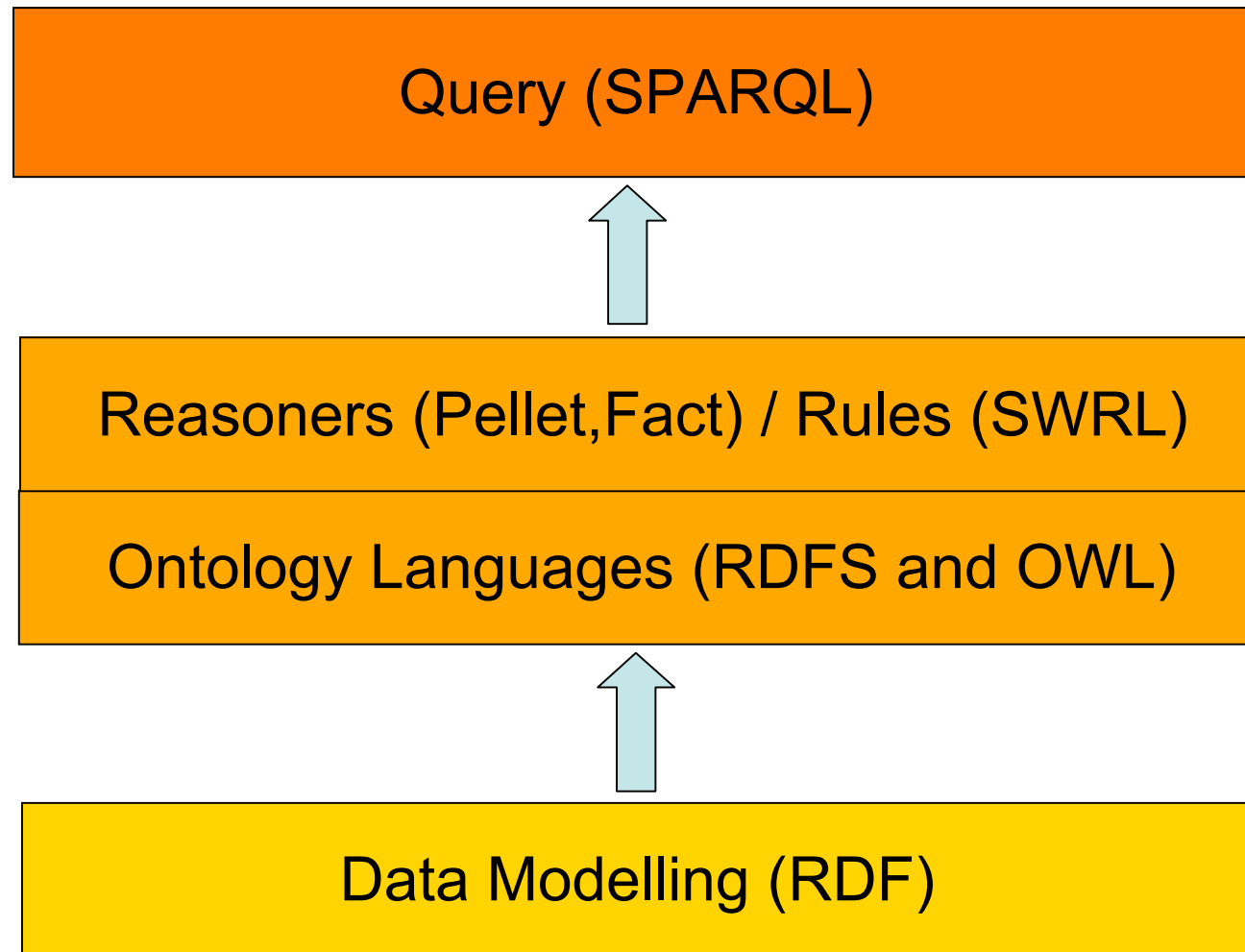
“An Ontology is a structure capturing semantic knowledge about a certain domain by describing relevant concepts and relations between them.” [5]

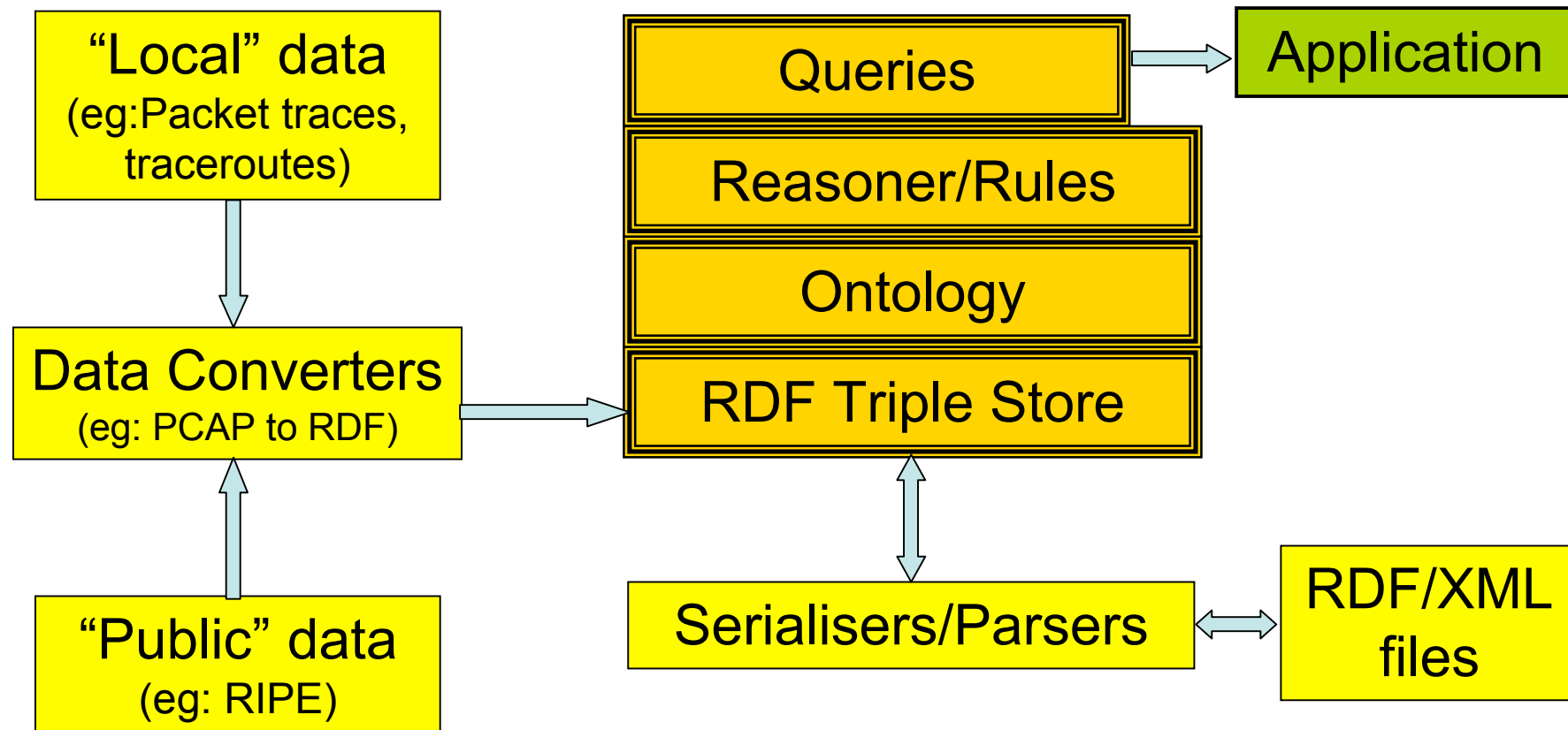


Layered Approach, built on the foundation of the Resource Description Framework (RDF) and the RDF 'triple' Statement :

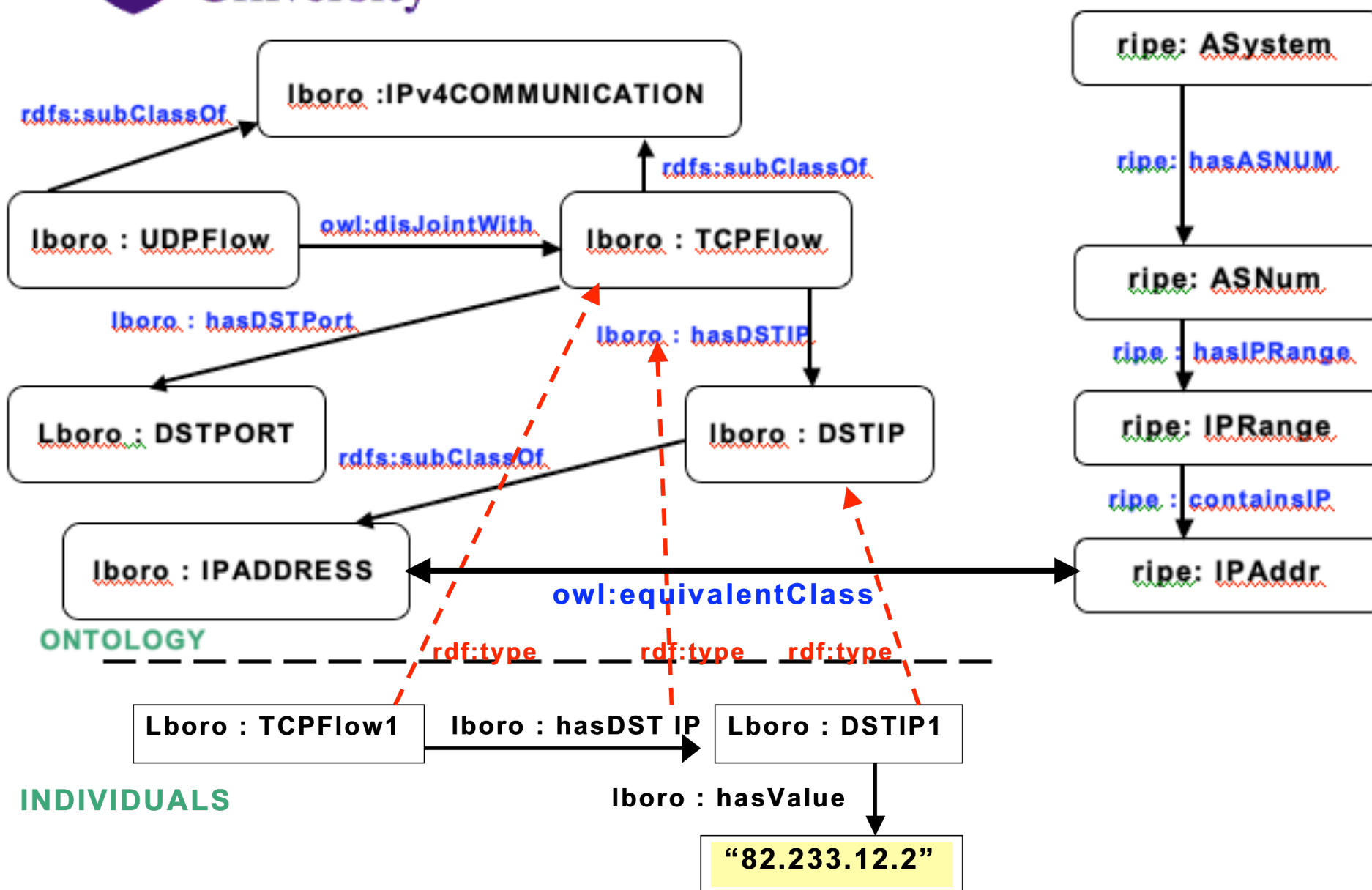


cos = "http://nets.lboro.ac.uk/home/coidn/cosenersontology.rdf/"





Sample Ontology : Graph snippet



```
<rdf:Description rdf:about="http://nets.lboro.ac.uk/TCPFLOW72">
  <lboro:hasRESET>false</lboro:hasRESET>
  <lboro:hasSRCIPAddr rdf:resource="http://nets.lboro.ac.uk/SRCIP72"/>
  <lboro:hasDSTPort>80</lboro:hasDSTPort>
  <rdf:type rdf:resource="http://nets.lboro.ac.uk/TCPFlow"/>
</rdf:Description>
—
<rdf:Description rdf:about="http://nets.lboro.ac.uk/UDPFlow">
  <owl:disjointWith rdf:resource="http://nets.lboro.ac.uk/TCPFlow"/>
  <rdfs:subClassOf rdf:resource="http://nets.lboro.ac.uk/IPCommunication"/>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
</rdf:Description>
```

- Expand ontology domain to include more data, specifically more converters.
- Develop richer semantics (predicates).
- Develop queries and application, liaising with industry.
- Initially concentrate on AS topology data and local packet captures.

Thank You. Any questions?

References

- [1] RIPE. www.ripe.net.
- [2] Colleen Shannon; David Moore; Ken Keys; Marina Fomenkov; Bradley Huffaker; K. Claffy. The internet measurement data catalog. ACM SIGCOMM Computer Communication Review, 35:97-100, 2005.
- [3] MOME. www.ist-mome.org.
- [4] Mahadevan; Krioukov; Fomenkov; Huffaker; Dimitropoulos; Claffy; Vahdat. The internet as-level topology: three data sources and one definitive metric. In ACM/SIGCOMM Computer Communications Review Vol 36 Jan 2006, 2006.
- [5] Marko Grobelnik; Dunja Mladenic. *Semantic Web Technologies: Trends and Research in Ontology Based Systems, chapter 2. Knowledge Discovery for Ontology Construction*. John Wiley and Sons, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, England, first edition, 2006.

Recommended Reading

Grigoris Antoniou; Frank van Harmelen. *A Semantic Web Primer*. The MIT Press, Cambridge, Massachusetts, second edition, 2008.

Dean Allemang; Jim Hendler. *Semantic Web for the Working Ontologist*, Elsevier Inc, Burlington, Massachusetts, first edition, 2007.

Related Work

<http://www.fp7-moment.eu/>

rdf

:type

rdfs

:subClassOf

:domain

:range

:subPropertyOf

owl

:class

:ObjectProperty

:datatypeProperty

:inverseOf

:SymmetricProperty

:ReflexiveProperty $\forall p \forall A \text{ for all } A$

:IrreflexiveProperty

:TransitiveProperty

- “Locally generated” data, including Passive packet captures at Internet gateways, Active traceroute data, etc.
- Regional Internet Registry data (eg: RIPE): delay measurements, AS to IP Address mappings etc [1].
- CAIDA: Many ongoing projects, including Archipelago (ARK) for AS level topology updates. Also CAIDA Internet Data Catalog [2].
- MoMe : EU FP6 initiative which gives a standardised interface to data from several ongoing projects and a data catalog [3].