Semantic e-Science: From Microformats to Models

L I Lumb, J R Freemantle & K D Aldridge

AGU 2009 Joint Assembly

IN24A-01

May 26, 2009

YORK
UNIVERSITÉ
UNIVERSITY

Semantic Web Hurdles

- Ontology writing
- Difficulty of implementation
- Competition
- Inaccuracy and deception

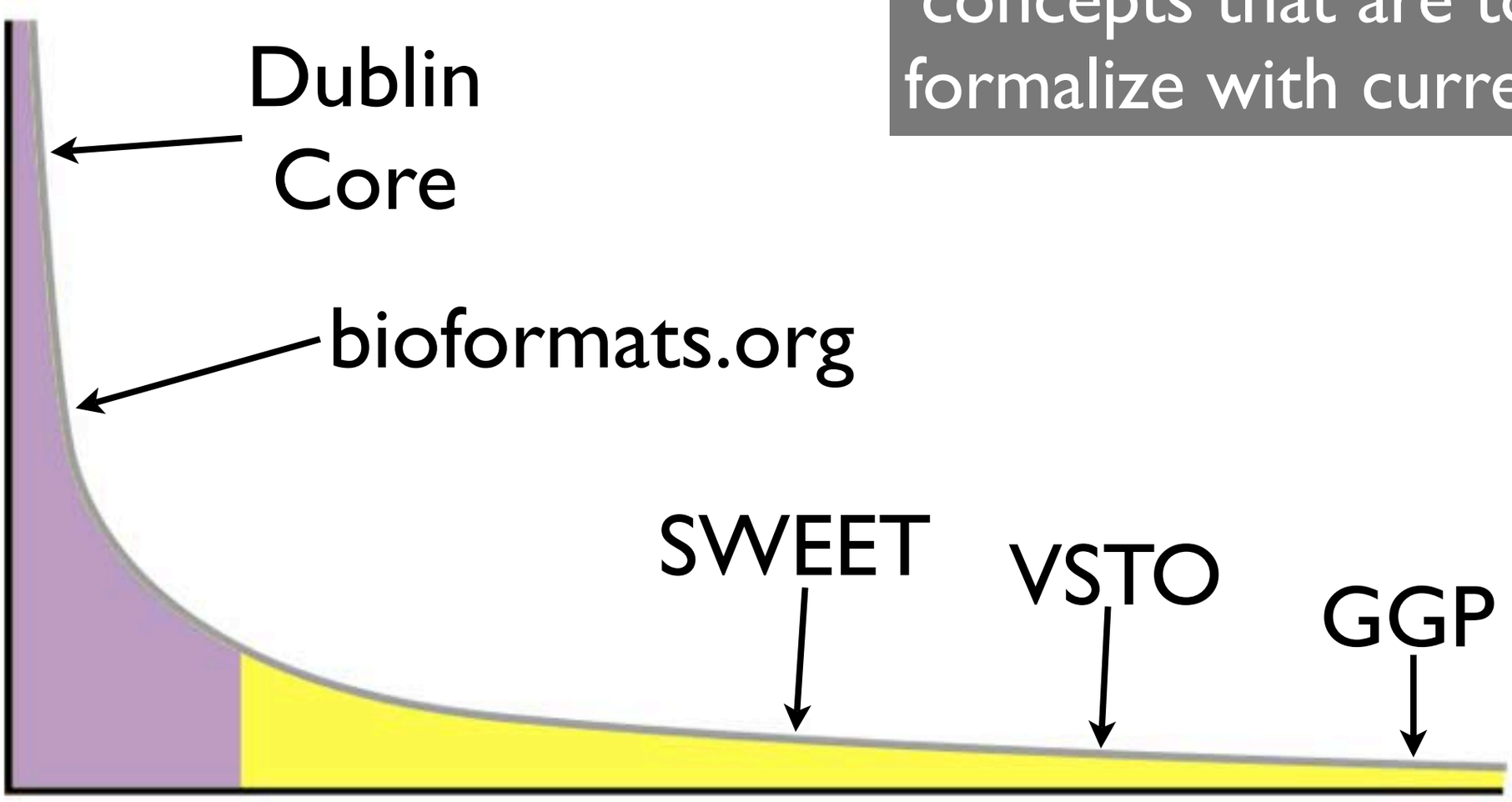
The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

IEEE INTELLIGENT SYSTEMS, 2009

Ontology Writing

"... a long tail of rarely used concepts that are too *expensive* to formalize with current technology."



Difficulty of Implementation

"Creating a database-backed Web service is substantially harder, requiring specialized skills. Making that service compliant with Semantic Web *protocols* is harder still."

Competition

http://en.wikiquote.org/wiki/Grace_Hopper

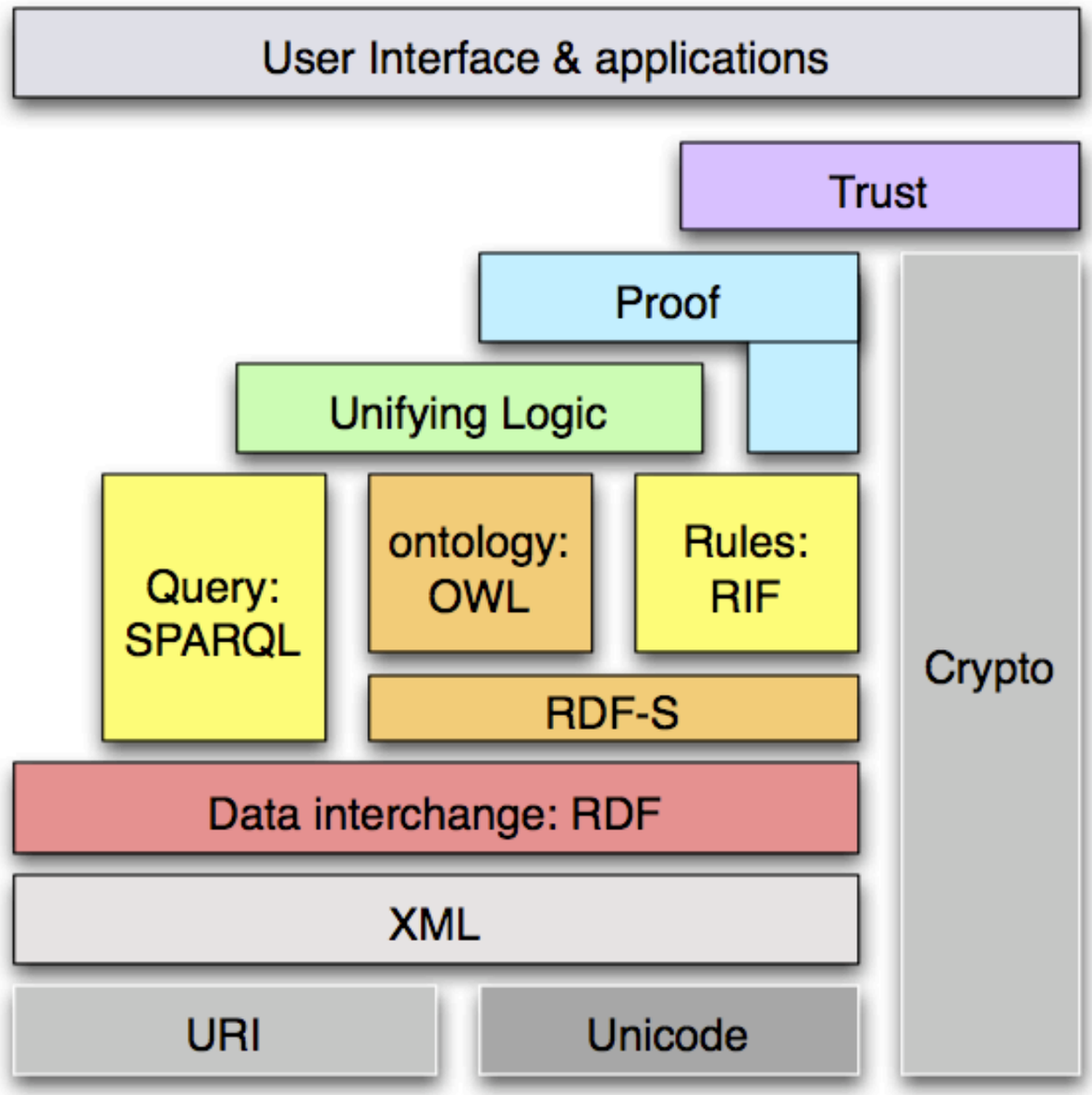
"The wonderful thing about *[ontologies]* is that there are so many of them to choose from."

Swoogle
semantic web search
2007

Inaccuracy and deception

"Anyone can say Anything about Any topic"

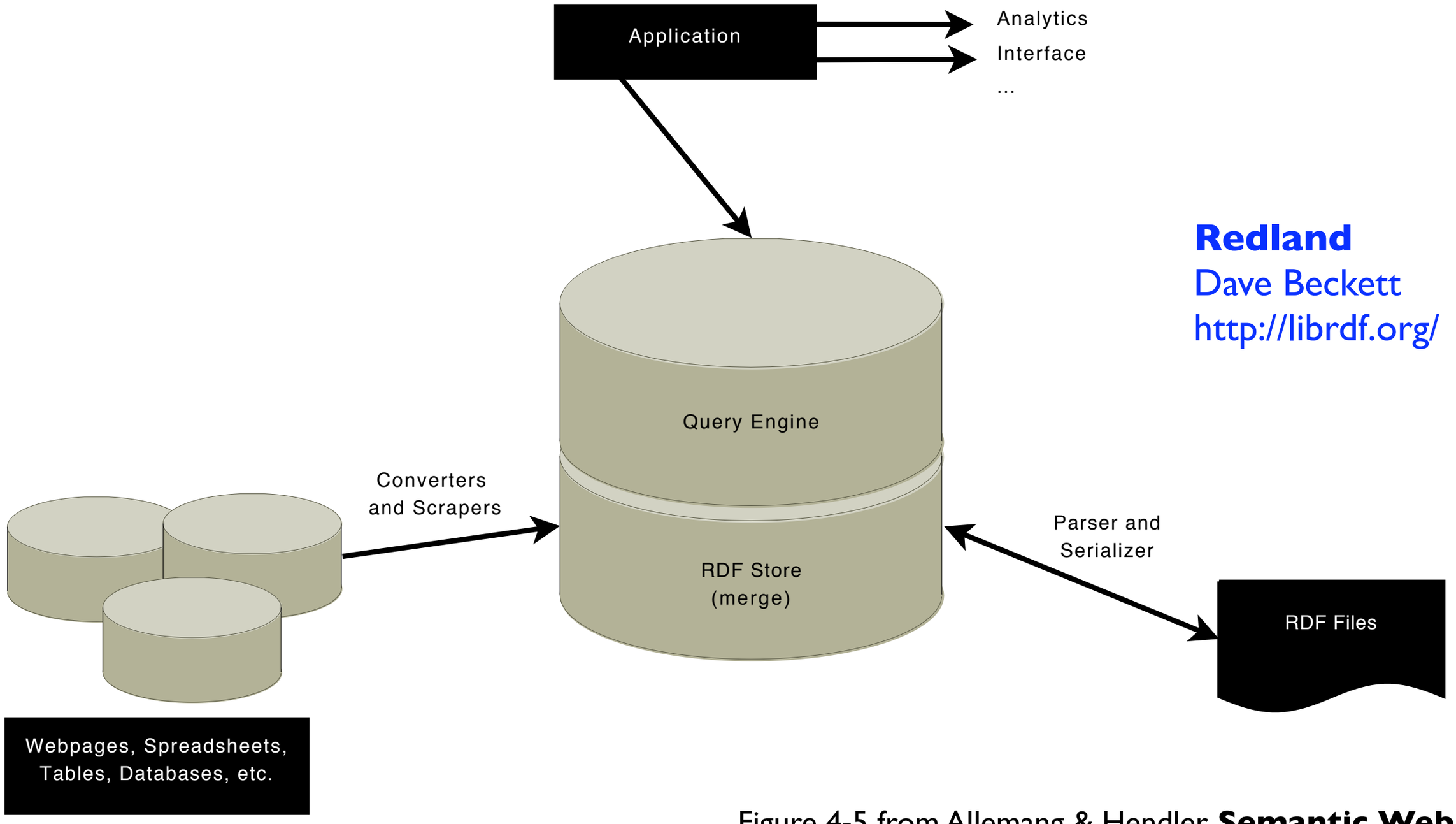
Berners-Lee's "Stack of Expressive Power"



Requirements

- Automate the process of 'writing' ontologies
- Automate integration of legacy databases with semantic platforms
- Automate integration of existing ontologies
- Establish a methodology for proof-based trust

RDF Application Architecture



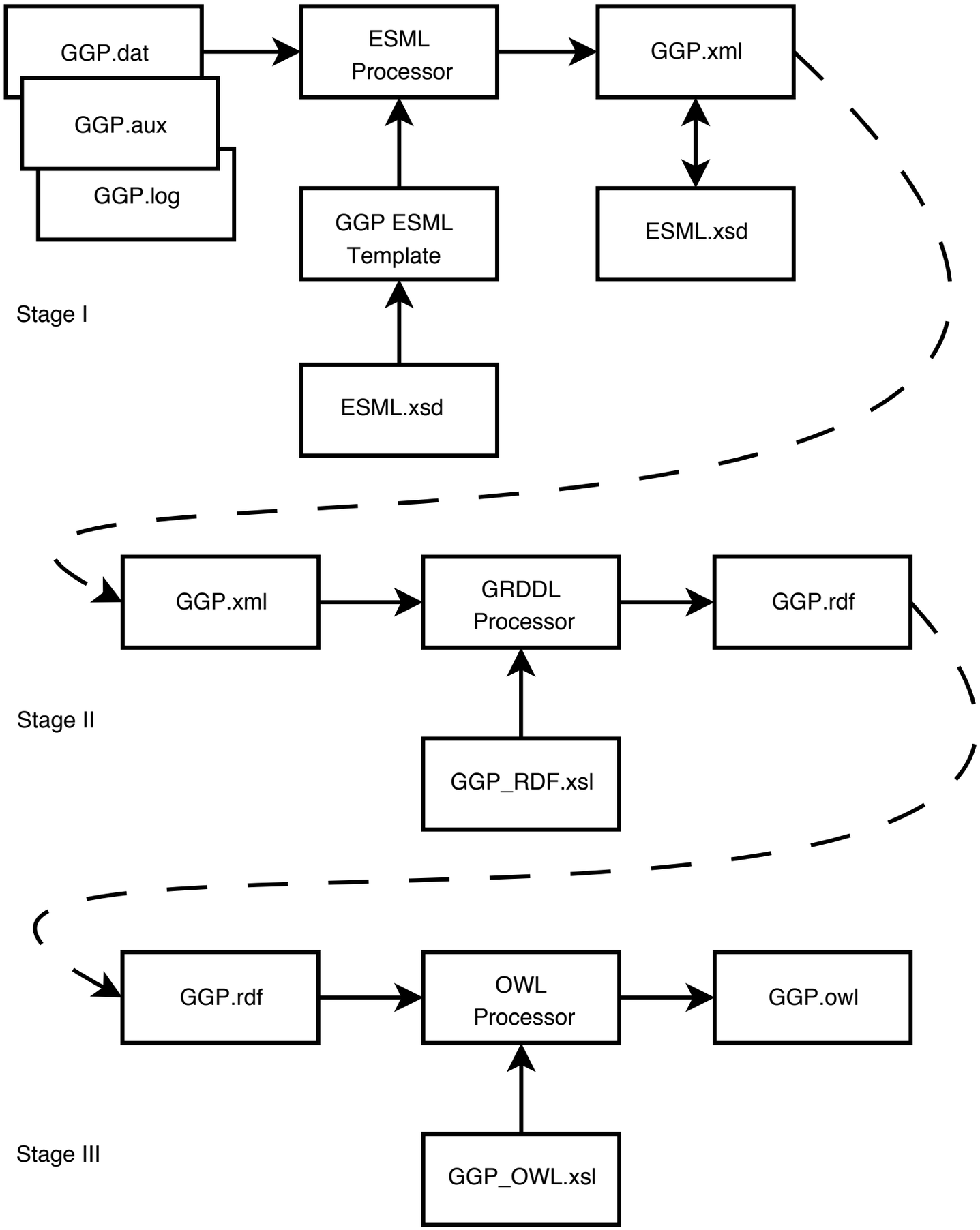
Redland
Dave Beckett
http://librdf.org/

Figure 4-5 from Allemang & Hendler, **Semantic Web for the Working Ontologist**, Morgan Kauffman, 2008

RDF Representation Algorithm

```
{Invoke external packages - e.g., RDF platform}
{Define variables - including the URIs for the required predicates}
{Instantiate a persistent database for this RDF representation}
{Instantiate an RDF model for this representation in the persistent
database}
while (executing SQL queries delivers data to standard output) do
   {Convert standard output into RDF subject-predicate-object triple}
   {Insert RDF triple as statement addition to RDF model}
end while
{Instantiate RDF/XML serializer}
{Establish namespaces for RDF/XML serialization}
{Serialize the RDF model to a file as output}
{Close all open files}
```

After Lumb, Gorsht & Zeng (submitted), **Data Management in Semantic Web**, Nova Science Publishers, Inc. http://grid.hust.edu.cn/DMSW-Book/



After Lumb, Gorsht & Zeng (submitted),
Data Management in Semantic Web,
Nova Science Publishers, Inc. http://grid.hust.edu.cn/DMSW-Book/

```
<?xml version="1.0" encoding="UTF-8"?>
<ESML xmlns="ESML" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<SyntacticMetaData>
 <Ascii>
  <Structure instances="1">
   <Array occurs="2">
    <Header name="_Instrument" format="%20s" />
    <Header name="Instrument" format="%20s" />
   </Array>
   <Array occurs="2">
    <Header name="_Latitude" format="%20s" />
    <Header name="Latitude" format="%10.4f" />
   </Array>
   <Array occurs="2">
    <Header name="_Longitude" format="%20s" />
    <Header name="Longitude" format="%10.4f" />
   </Array>
   <Array occurs="2">
    <Header name="_Height" format="%20s" />
    <Header name="Height" format="%10.4f" />
   </Array>
  </Structure>
 </Ascii>
</SyntacticMetaData>
</ESML>
```

CLASS

PROPERTY

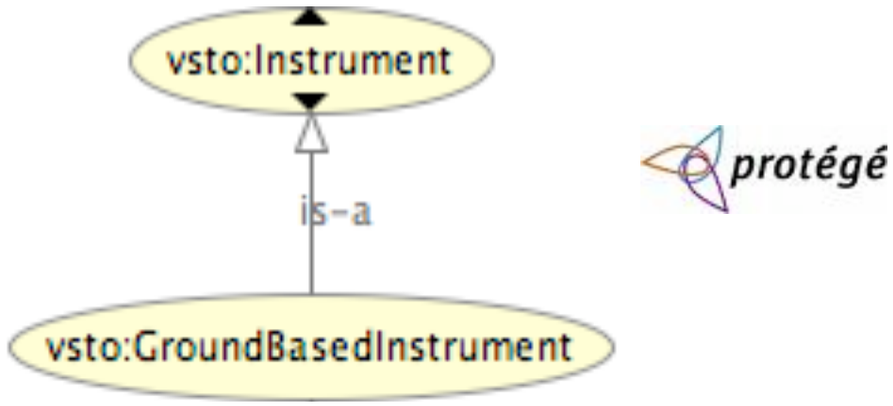
PROPERTY

PROPERTY

ESML Template for the GGP

Lumb & Aldridge (HPCS 2005)

http://esml.itsc.uah.edu

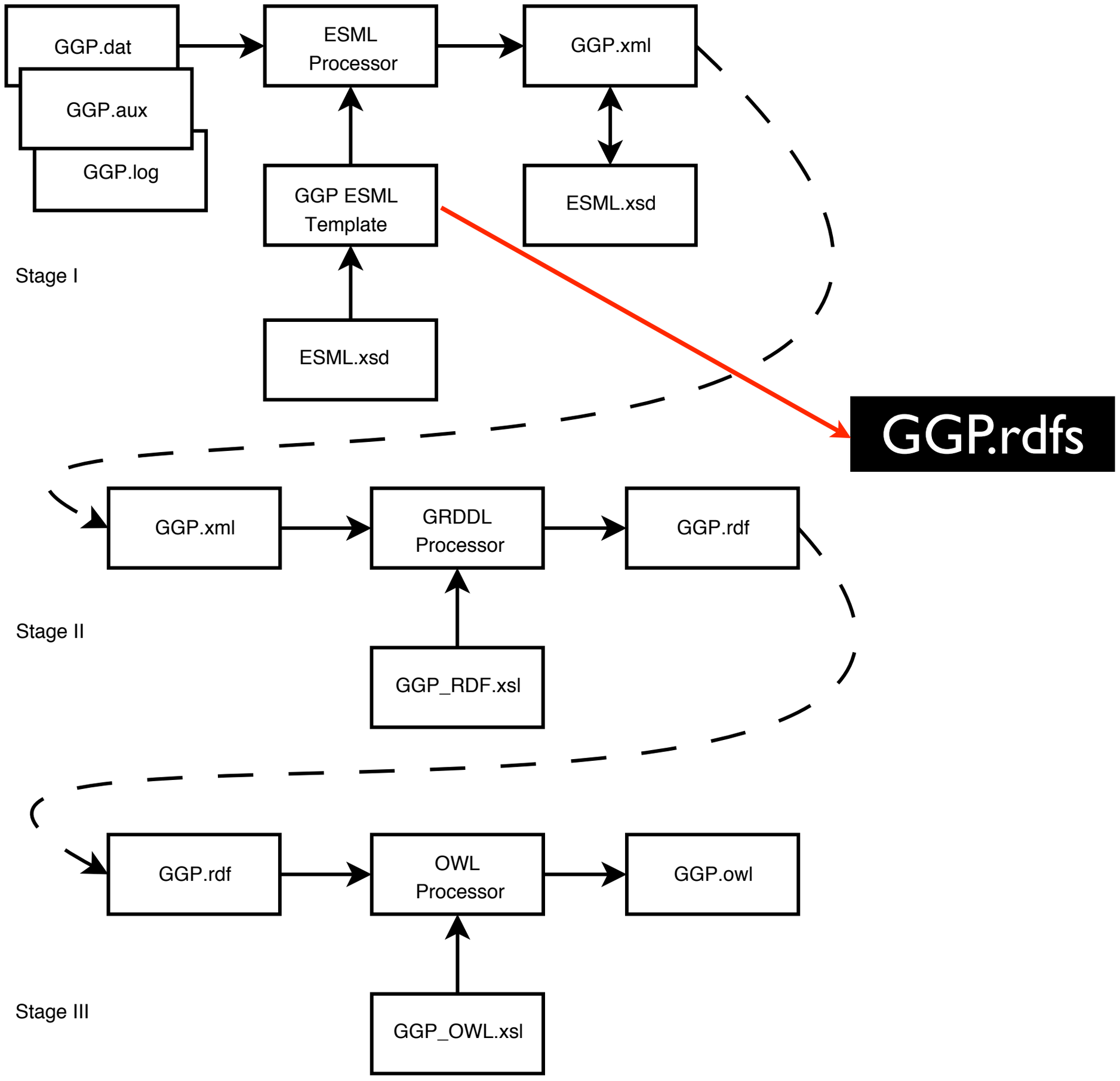


owl:equivalentClass

Instrument

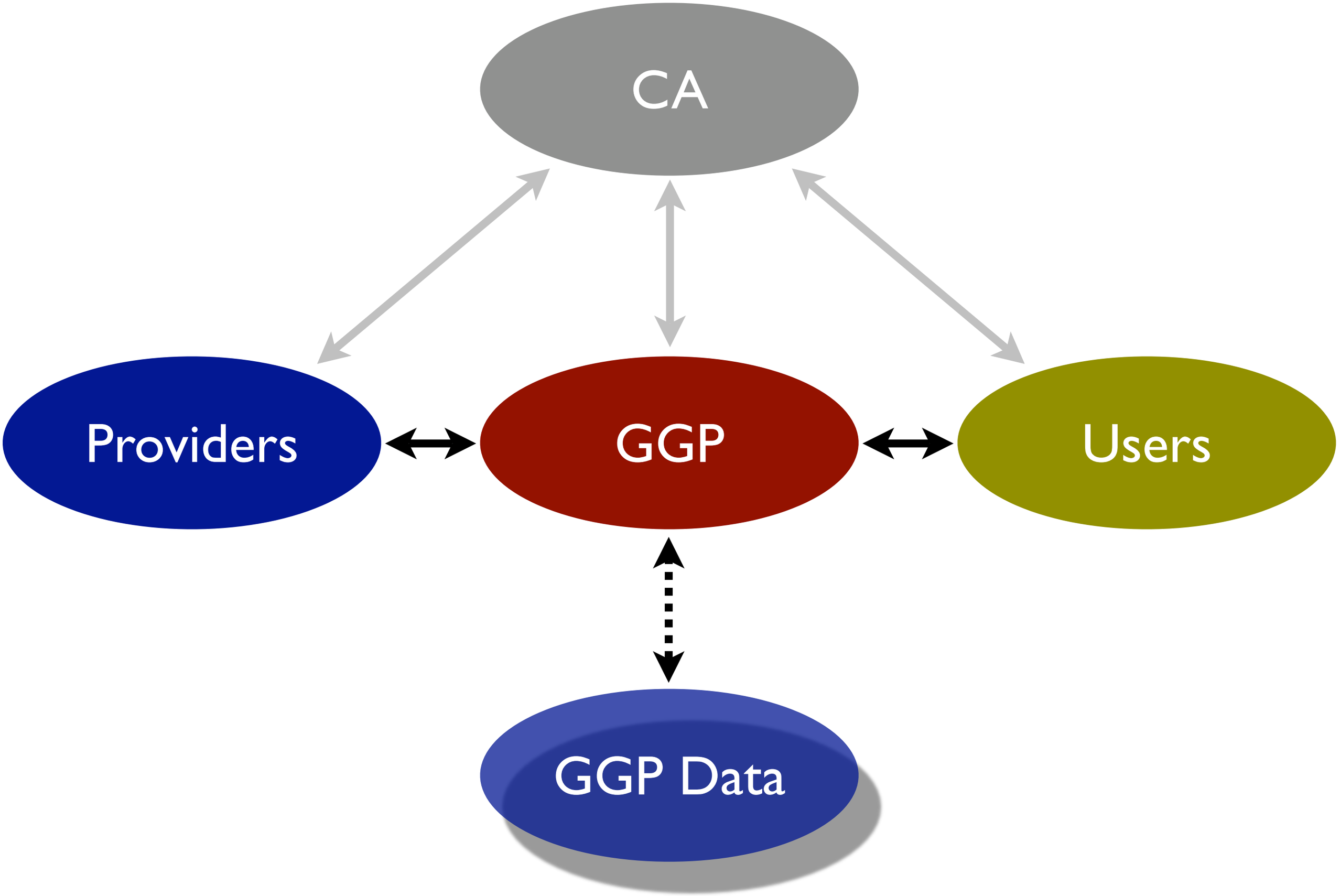
(RDFS-Plus)

Lumb et al., *Comp & Geosci.*,
35, 855-861, 2009


GGP.dat
GGP.aux
GGP.log
ESML Processor
GGP.xml
GGP ESML Template
ESML.xsd
ESML.xsd
Stage I
GGP.rdfs
GGP.xml
GRDDL Processor
GGP.rdf
Stage II
GGP_RDF.xsl
GGP.rdf
OWL Processor
GGP.owl
Stage III
GGP_OWL.xsl

Summary

- Automate the process of 'writing' ontologies
- Automate integration of legacy databases with semantic platforms
- Automate integration of existing ontologies
- Establish a methodology for proof-based trust


CA
Providers
GGP
Users
GGP Data

Additional Slides

Query Existing Database Algorithm

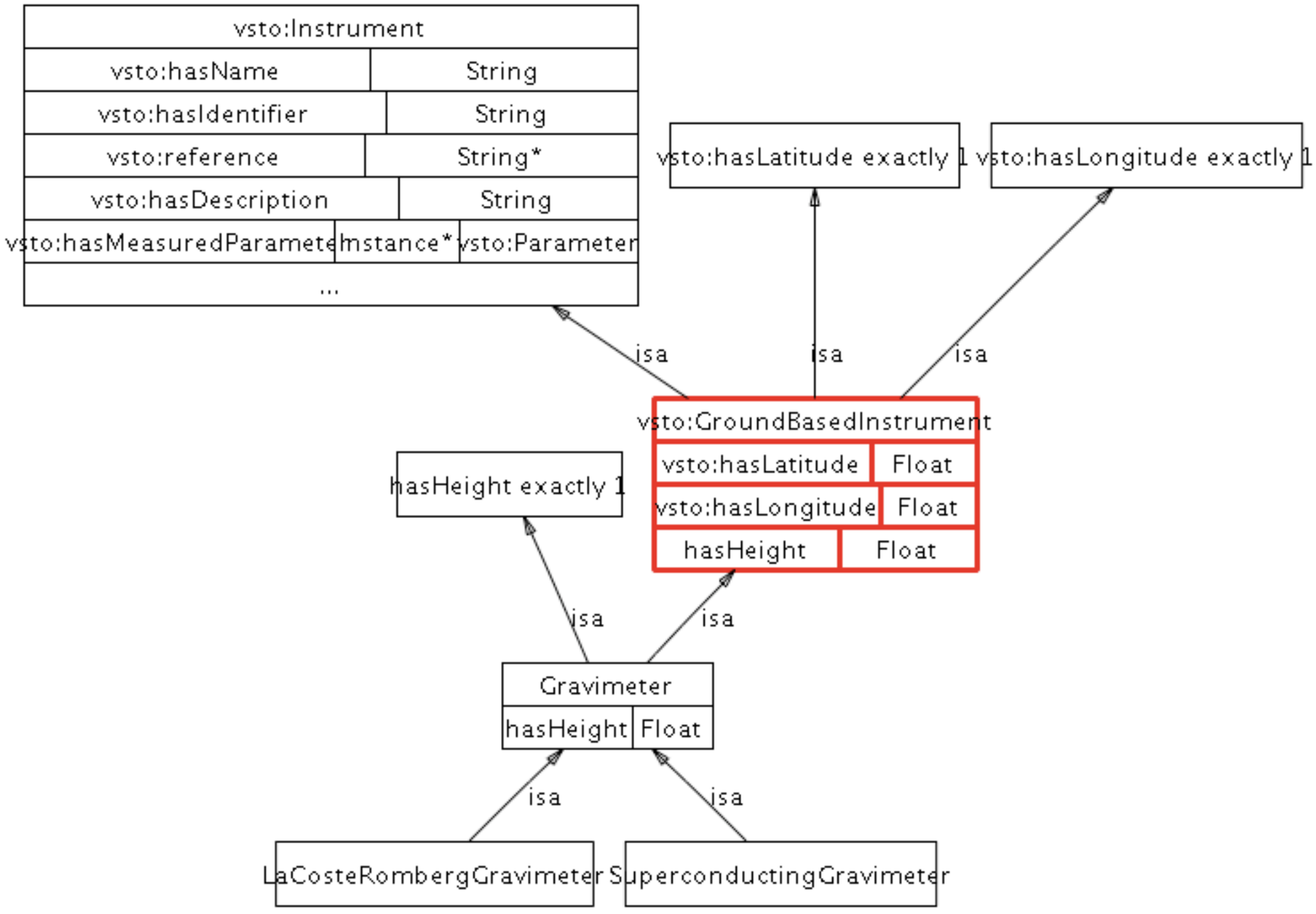
```
{Invoke external packages - e.g., to facilitate interaction with databases}
{Accept input (if necessary)}
{Define variables}
{Define database specifics - including server and port for access}
{Open connection with the database}
{Enter query in SQL to extract data of interest}
select ...
{Execute SQL query}
while (the SQL query returns results) do
   {Extract additional details via additional SQL queries}
   {Print query results to standard output}
end while
{Close connection with database}
```

After Lumb, Gorsht & Zeng (submitted), **Data Management in Semantic Web**, Nova Science Publishers, Inc. http://grid.hust.edu.cn/DMSW-Book/

RDF Querying Algorithm

```
{Invoke external packages - e.g., RDF platform}
{Define variables - including the name of the .rdf file}
{Open the .rdf file for input}
{Instantiate a persistent database for this RDF representation}
{Instantiate an RDF model for this representation in the persistent database}
{Instantiate RDF/XML parser}
while (there is data from the file to parse as a stream) do
   {Add parsed streams as statements to the RDF model}
   {Detect namespaces during parsing and store to an array}
end while
{Open the file containing the query in SPARQL for reading}
while (there are statements in the RDF model to search against as a stream) do
   {Query each statement for search-term match}
   {Save matches in an array}
end while
{Instantiate RDF/XML serializer}
{Establish namespaces for RDF/XML serialization}
{Serialize query results to a .rdf file as output}
{Close all open files}
{Deconstruct the serializer, persistent data store and RDF model}
{Visualize results in the .rdf file graphically}
```

After Lumb, Gorsht & Zeng (submitted), **Data Management in Semantic Web**, Nova Science Publishers, Inc. http://grid.hust.edu.cn/DMSW-Book/


vsto:Instrument
vsto:hasName
String
vsto:hasIdentifier
String
vsto:reference
String*
vsto:hasDescription
String
vsto:hasMeasuredParameter
Instance*
vsto:Parameter
...
vsto:hasLatitude exactly 1
vsto:hasLongitude exactly 1
isa
isa
isa
vsto:GroundBasedInstrument
vsto:hasLatitude
Float
vsto:hasLongitude
Float
hasHeight
Float
hasHeight exactly 1
isa
isa
Gravimeter
hasHeight
Float
isa
isa
LaCosteRombergGravimeter
SuperconductingGravimeter

Original Motivation

- Scientific
  - Temporal and/or spatial alignment of GGP data is a manual process
  - Complex GGP data reductions are required
  - Numerous unwanted signals need to be removed
- Computer Science
  - Applicability of Grid Computing in the case of small-to-medium-scale science