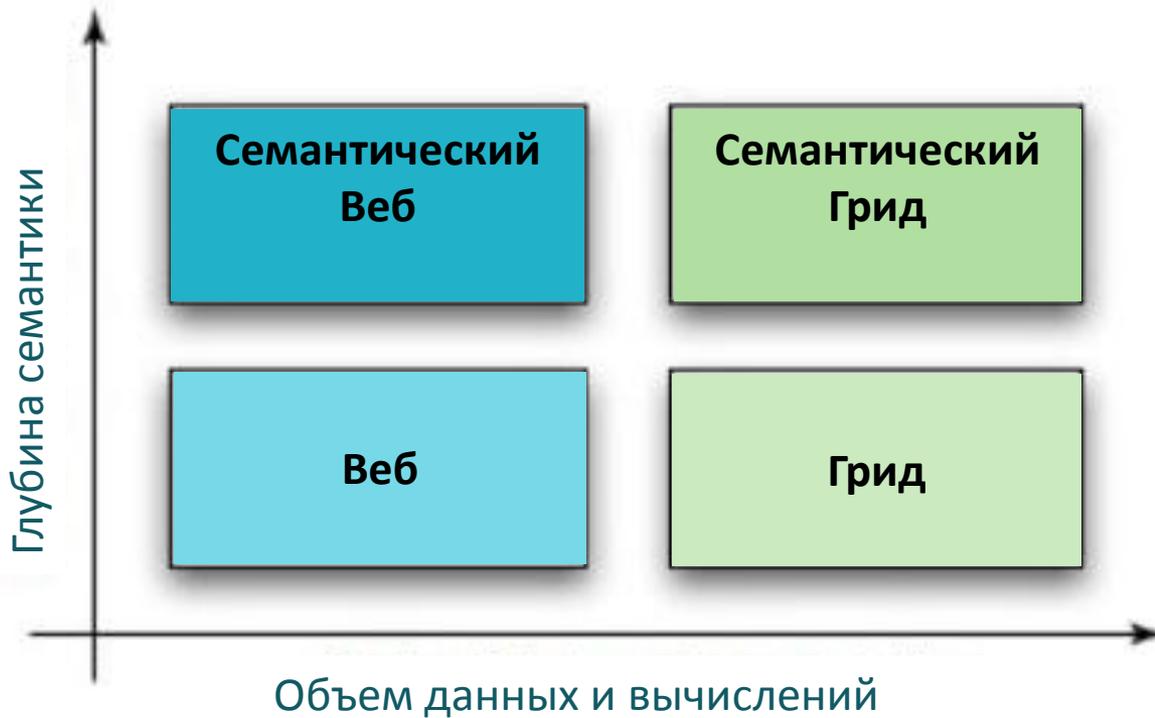


# Данные и метаданные в Семантическом Грид

Дрозд А.Ю., НТУУ «КПИ», ИПСА

# Семантический Грид



# Семантический Грид

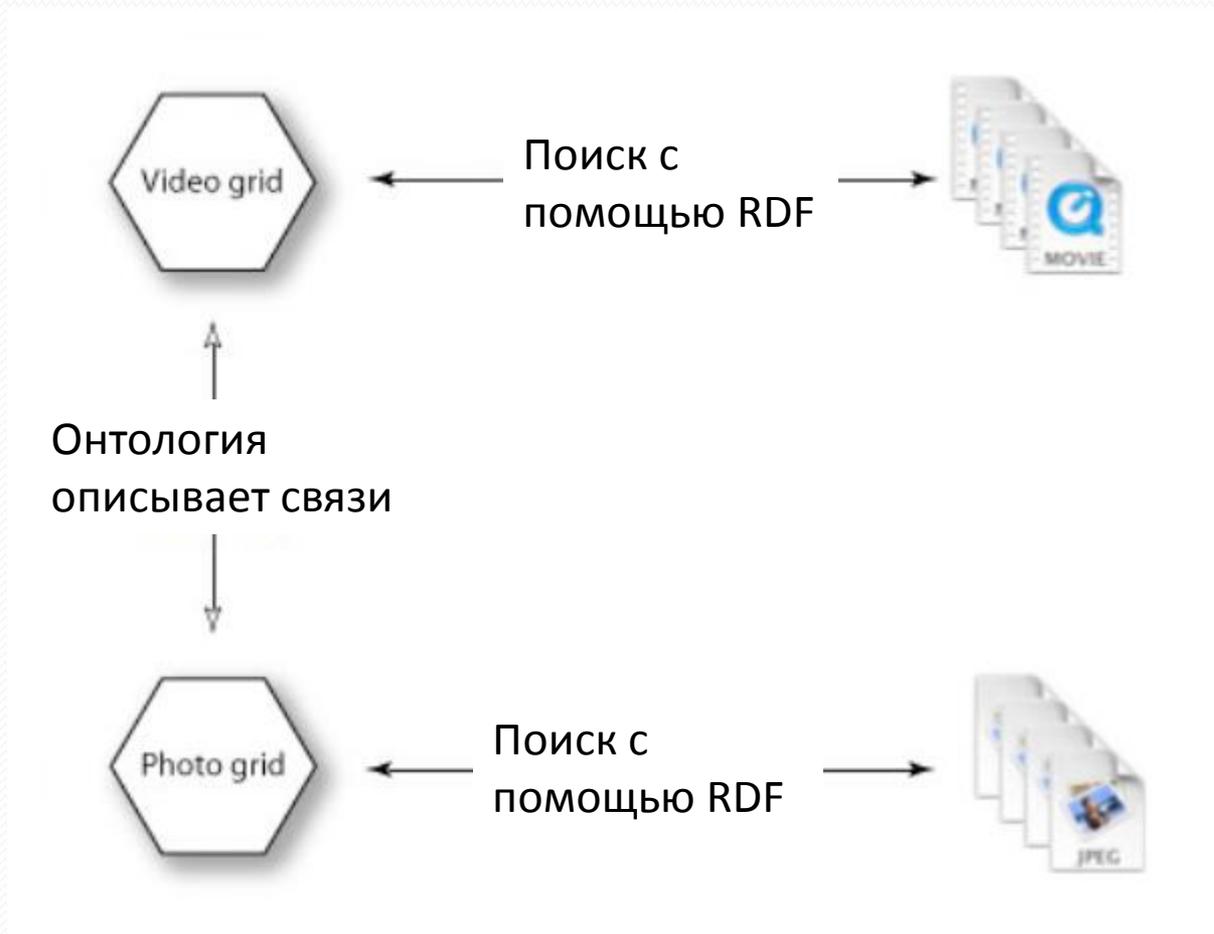
- *The Semantic Grid is an extension of the current Grid in which information and services are given well-defined meaning through machine-processable descriptions which maximize the potential for sharing and reuse.*

*David De Roure*

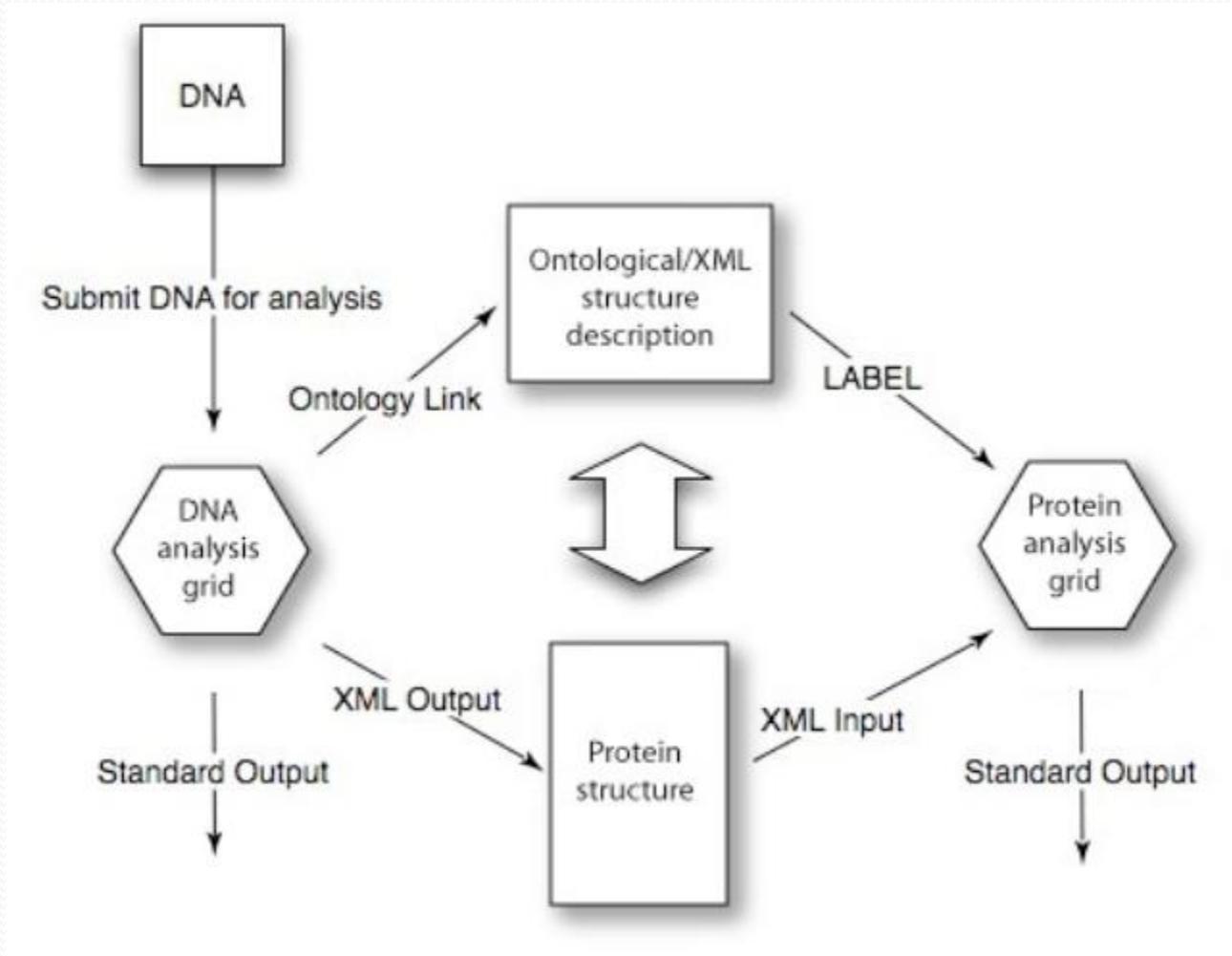
- Семантический Грид это такое расширение существующего Грид, в котором информации и сервисам дано четко-определенное значение посредством машинно-обрабатываемых описаний для максимального их использования.

Девид Де Рур

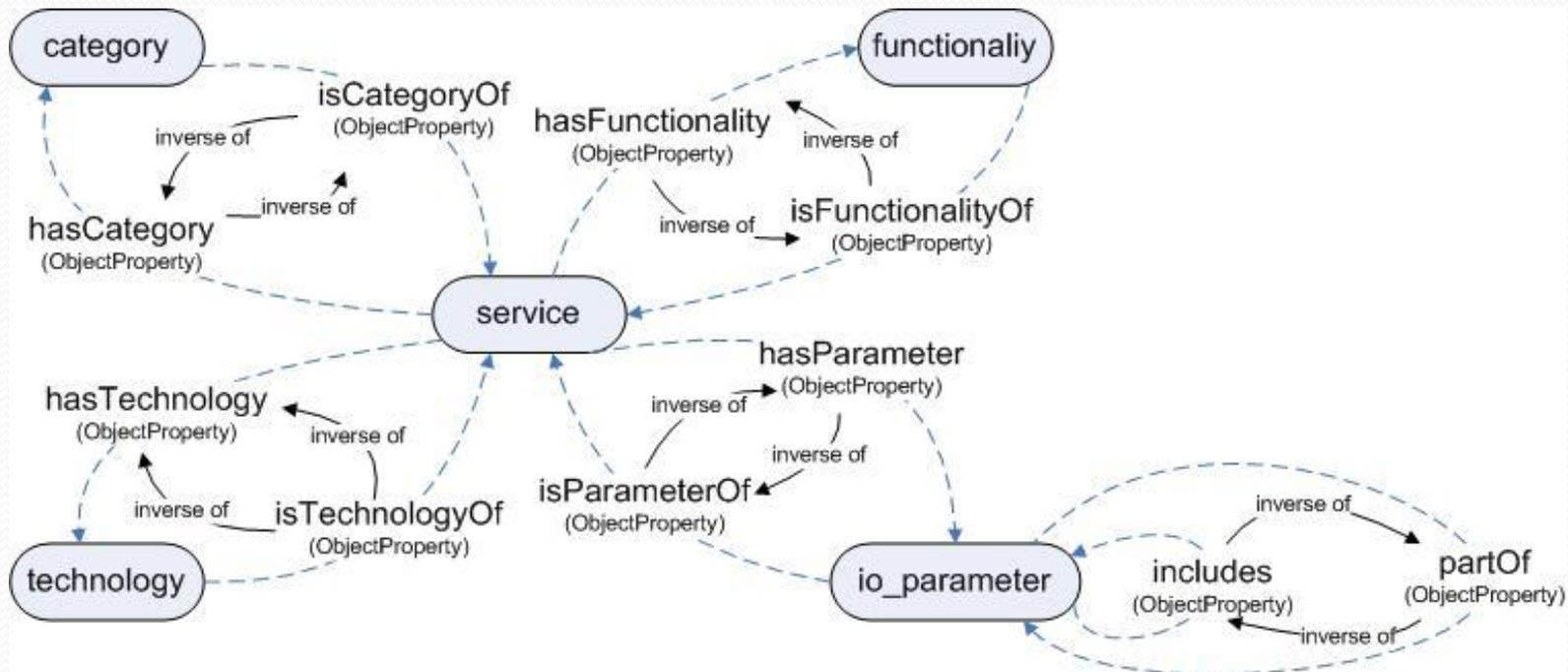
# Связи между данными



# Связи между ресурсами



# Связи между сервисами



# Метаданные

- *Метаданные* — это информация об используемых данных
- Структурированные данные, представляющие собой характеристики описываемых сущностей для целей их идентификации, поиска, оценки, управления ими
- Небольшие записи высокой ценности

# Типы метаданных

- Регистры, представляющие URL(адреса) в форме URI(уникальные идентификаторы) для каждого Грид ресурса или сервиса;
- *Для ресурсов:* метаданные, характеризующие базы данных, пользователей, политики, квоты, заполненность и т.п.
- *Для сервисов:* базовые списки доступных сервисов с заданной функциональностью и расположением, без углубленного контекста;
- *Для Семантического Грид в целом:* сложные онтологии и метаданные для интеллектуального поиска и сопоставления сервисов и ресурсов.

# Типы метаданных

- *С точки зрения мониторинга:* потоки информации, отображающие выполнение операций Грид-сервисами. Среди них: данные по статусам передачи по сети и выполнения задач;
- *С точки зрения сотрудничества:* данные по изменению состояний сервисов и ресурсов для поддержания соответствия между разными версиями и копиями одинаковых базовых ресурсов;
- *С точки зрения каждой прикладной области:* специализированные метаданные для управления данными в каждой из областей. Эти метаданные должны храниться в специальных каталогах, которые обслуживаются в форматах, соответствующих каждой прикладной области;
- *С точки зрения происхождения и управления данными:* метаданные, описывающие жизненный цикл и права собственности на данные. Эта область, соединяющая в себе Грид и права на интеллектуальную собственность, требует особого внимания.

# Форматы метаданных



Глубина

семантики

**XML** предоставляет синтаксис для определения структуры документа, подлежащего машинной обработке. Не несёт семантической нагрузки.

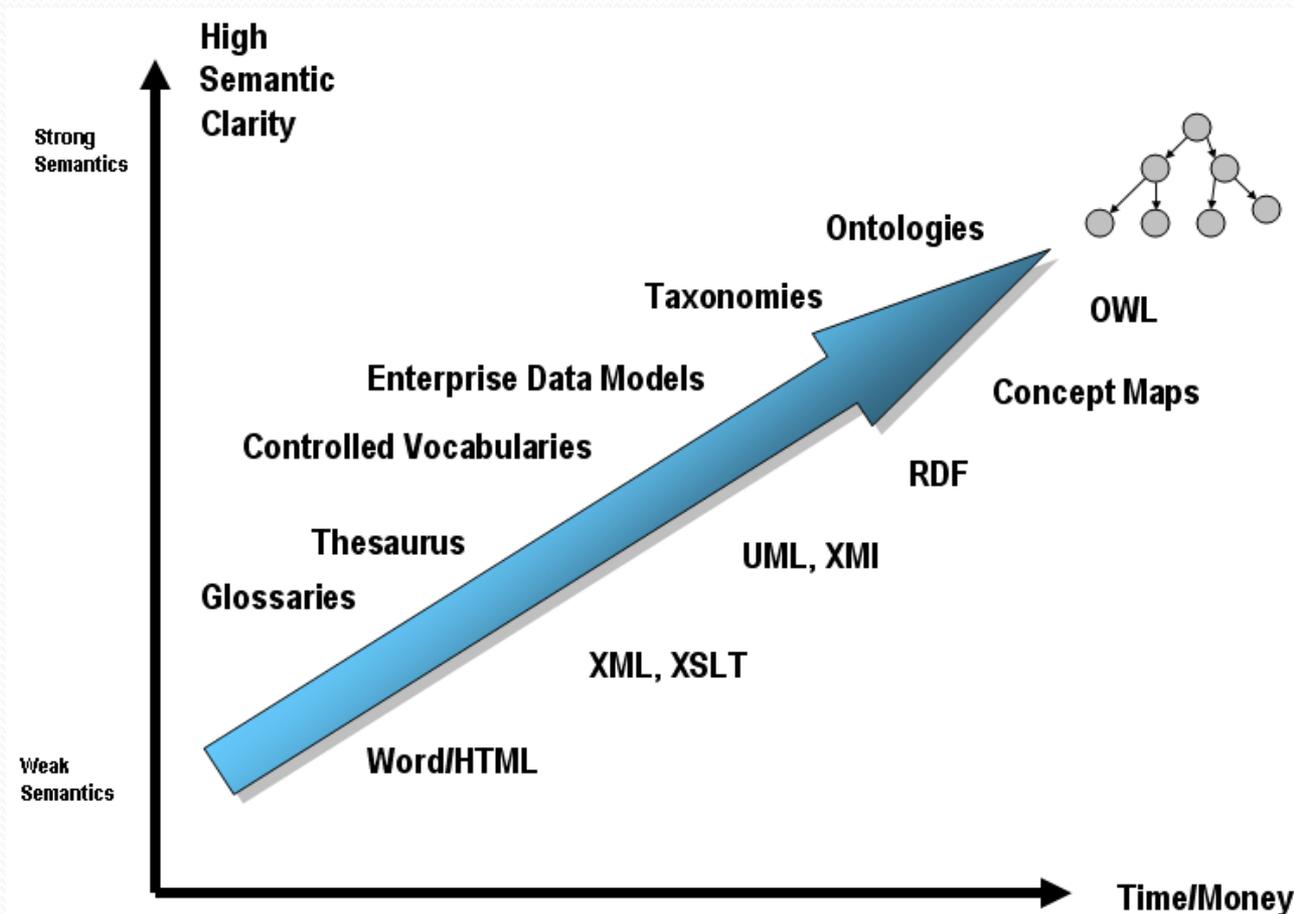
**XML Schema** определяет ограничения на структуру XML-документа.

**RDF** представляет собой простой способ описания экземплярных данных в формате **субъект-отношение-объект**, в котором в качестве любого элемента этой тройки используются только идентификаторы ресурсов. Существует стандартизованное отображение этих троек на XML-документы.

**RDF Schema** описывает набор атрибутов (здесь их точнее назвать отношениями), таких, как `rdfs:Class`, для определения новых типов RDF-данных. Языком поддерживается также отношение наследования типов `rdfs:subClassOf`.

**OWL** расширяет возможности по описанию новых типов (в частности, добавлением перечислений), а также позволяет описывать новые типы данных **RDF Schema** в терминах уже существующих (например, определять тип, являющийся пересечением или объединением двух существующих).

# Развитие метаданных



# Фрагмент RDF документа

```
@gridBackground { // specifies grid background knowledge
  Linux[rdfs:subClassOf->GR:OperatingSystem].
  Unix[rdfs:subClassOf->GR:OperatingSystem].
  Debian[rdf:type->Linux].   Redhat[rdf:type->Linux].
  SunOS[rdf:type->Unix].     Linux[rdf:type->Unix].

  // transitivity axiom
  FORALL X,Y,Z X[compatibleWith->Z] <- X[compatibleWith->Y] AND Y[compatibleWith->Z].

  // identity axiom
  FORALL X X[compatibleWith->X].

  //symmetry axiom
  FORALL X,Y X[compatibleWith->Y] <- Y[compatibleWith->X].

  FORALL X,Y,Z X[substitutes->Z] <- (Y[rdf:type->Z] and
    X[substitutes->Y]) or X[compatibleWith->Z].
}
```

# Фрагмент онтологии OWL

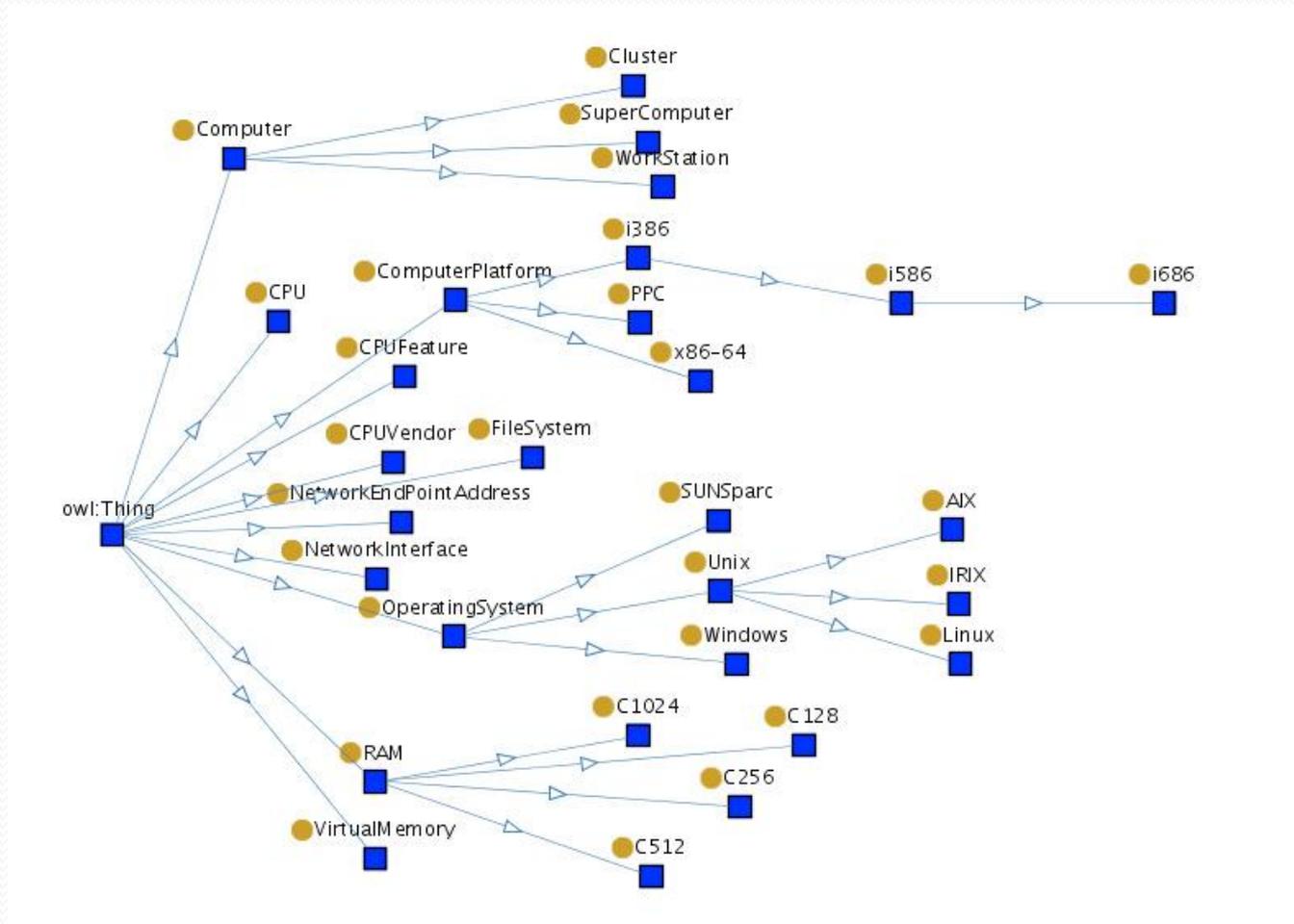
```
- <rdfs:domain>
  <owl:Class rdf:about="file:/C:/Program%
    20Files/OilEd/ontologies/ESG/jdl_example.daml#dataset" />
</rdfs:domain>
- <rdfs:range>
  - <owl:Class>
    - <owl:oneOf>
      - <rdf:List>
        - <rdf:first>
          <owl:Thing rdf:about="file:/C:/Program%
            20Files/OilEd/ontologies/ESG/jdl_example.daml#cloud_medium" />
          </rdf:first>
        - <rdf:rest>
          - <rdf:List>
            - <rdf:first>
              <owl:Thing rdf:about="file:/C:/Program%
                20Files/OilEd/ontologies/ESG/jdl_example.daml#bounds_latitude" />
              </rdf:first>
            - <rdf:rest>
              - <rdf:List>
                - <rdf:first>
                  <owl:Thing rdf:about="file:/C:/Program%
                    20Files/OilEd/ontologies/ESG/jdl_example.daml#temperature" />
                  </rdf:first>
                <rdf:rest rdf:resource="http://www.w3.org/2002/07/owl#nil" />
                </rdf:List>
              </rdf:rest>
            </rdf:List>
          </rdf:rest>
        </rdf:List>
      </owl:oneOf>
    </owl:Class>
  </rdfs:range>
</owl:ObjectProperty>
+ <owl:ObjectProperty rdf:about="#jdl_example.daml#hasParameter">
- <rdf:Description rdf:about="file:/C:/Program%
  20Files/OilEd/ontologies/ESG/jdl_example.daml#temperature">
  - <rdfs:comment>
    <![CDATA[   ]]>
  </rdfs:comment>
  - <oiled:creationDate>
    <![CDATA[ 2003-12-09T21:27:14Z   ]]>
  </oiled:creationDate>
  - <oiled:creator>
    <![CDATA[ 71p   ]]>
  </oiled:creator>
  - <rdf:type>
    <owl:Class rdf:about="file:/C:/Program%
      20Files/OilEd/ontologies/ESG/jdl_example.daml#parameter" />
  </rdf:type>
</rdf:Description>
```

Parameter:  
Cloud\_medium

Parameter:  
Bounds\_latitude

Parameter:  
temperature

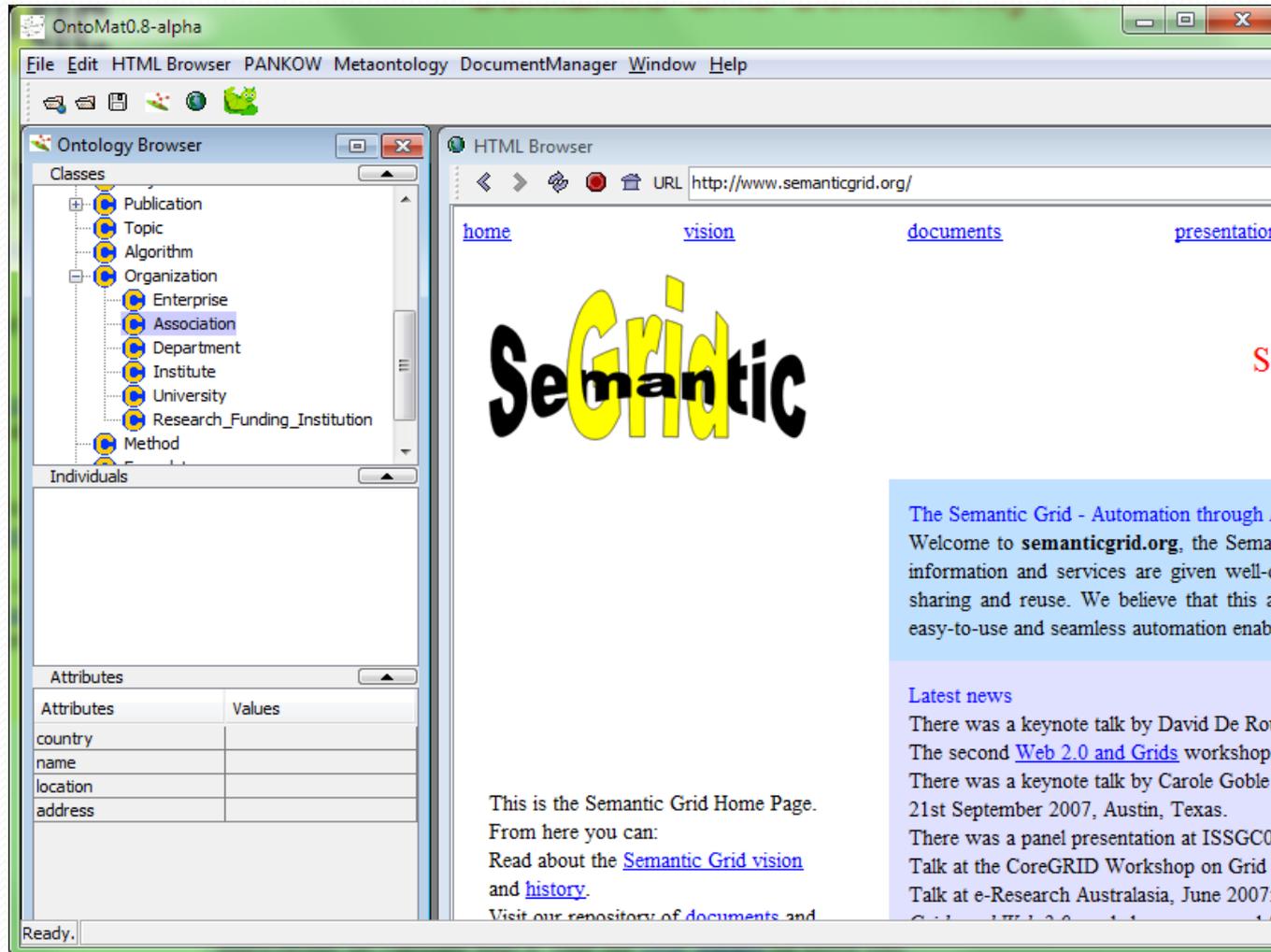
# Пример онтологии



# Генерирование метаданных

- **Ручное аннотирование:** в каждой прикладной области специалисты с достаточной квалификацией добавляют семантические метаданные к имеющимся ресурсам. Существуют отдельные организации, задача которых – семантическое аннотирование.
- **Автоматическое аннотирование:** так как метаданные машинно-читаемы, то на основе уже имеющихся данных и метаданных могут быть сгенерированы новые метаданные.  
*Автоматическое получение знаний(данные+метаданные) из существующих – одна из важнейших и сложнейших задач, которые еще предстоит решить.*

# OntoMat

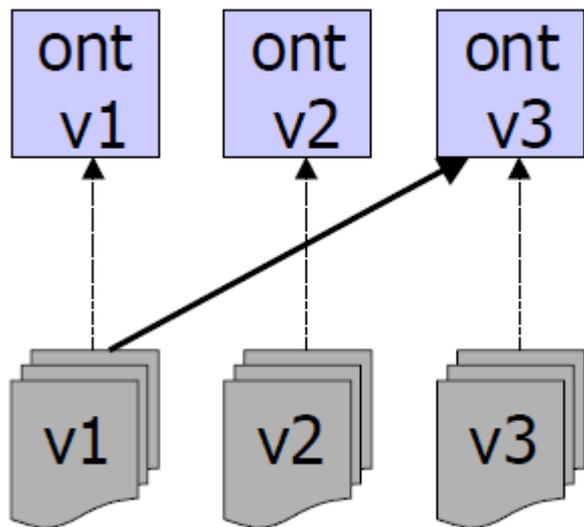


# Эволюция метаданных

- Семантический Грид – динамичная структура
- Метаданные постоянно изменяются
- Невозможно синхронизировать
- Необходим механизм присвоения **версий**

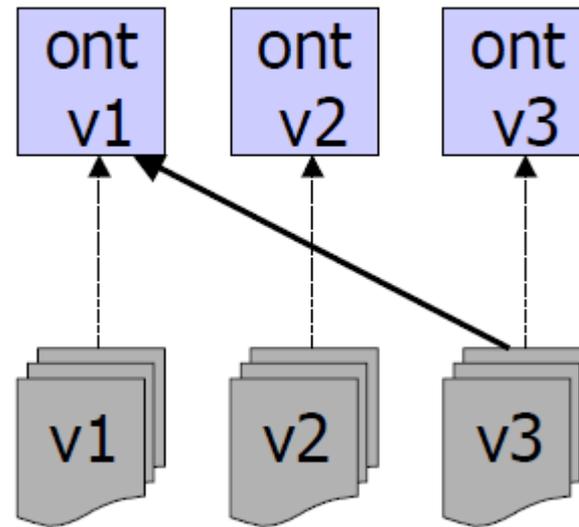
# Версии онтологий

## Перспективные версии



- Старые данные совместимы с новыми онтологиями

## Ретроспективные версии



- Новые данные совместимы со старыми онтологиями

# Совместимость версий

- Несовместимые версии – семантика новых версий онтологий изменена таким образом, что невозможно корректно интерпретировать данные предыдущих версий
- Прямая совместимость – новые версии данных могут быть корректно интерпретированы старыми онтологиями
- Обратная совместимость – старые версии данных могут быть корректно интерпретированы новыми онтологиями
- Полная совместимость – обе комбинации онтологий и данных дают корректный результат

