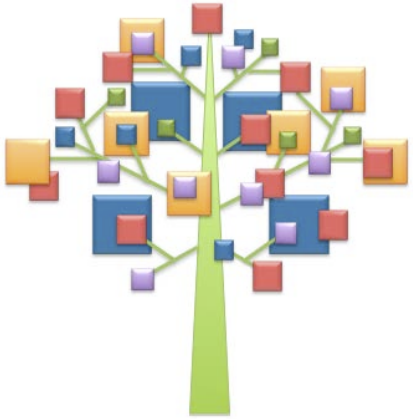


Ontology 101: An Introduction to Knowledge Representation & Ontology Development

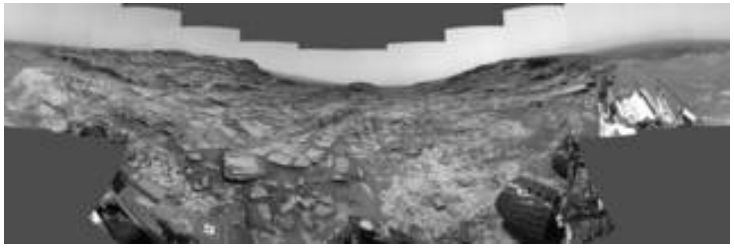
Elisa Kendall, Thematix Partners LLC & Deborah McGuinness, RPI / McGuinness Associates

18 August 2015

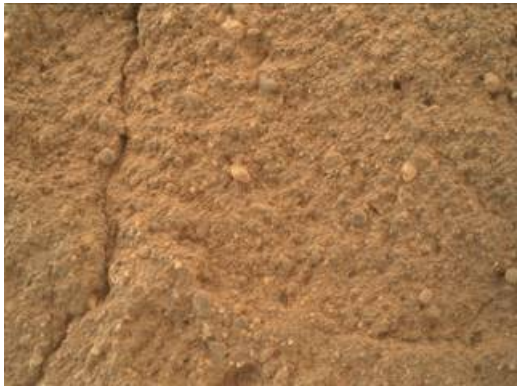
Content management

Mars was photographed by the Hubble Space Telescope in August 2003 as the planet passed closer to Earth than it had in nearly 60,000 years. Image Credit: NASA, J. Bell (Cornell U.) and M. Wolff (SSI)

A sunset on Mars creates a glow due to the presence of tiny dust particles in the atmosphere. This photo is a combination of four images taken by Mars Pathfinder, which landed on Mars in 1997. Image credit: NASA/JPL



This 360-degree panorama from NASA's Curiosity Mars rover shows the surroundings of a site on lower Mount Sharp where the rover spent its 1,000th Martian day, or sol, on Mars, in May 2015. Image credit: NASA/JPL



This May 29, 2015, view of a Martian sandstone target called "Big Arm" covers an area about 1.3 inches wide in detail that shows differing shapes and colors of sand grains in the stone. Image credit: NASA/JPL

The Planetary Data Store (PDS) is a distributed repository of 40+ years' imagery & data taken by a range of instruments on many diverse missions, available for scientific research.

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Smart search

Family Tree | Pedigree | Family | Family Group Sheet | Print

- Robbie M Hinkle 1907-1997
  - Clarence H Finnie
    - Clarence H Finnie
    - Robbie Lea Finnie
    - Ima Gene Finnie
    - + Add Child
  - + Add Spouse
  - Show siblings
  - Jefferson Hinkle 1881-1952
    - Joshua W C Hinkle 1848-1929
      - John D Hinkle 1816-1859
        - William Hinkle
        - Sarah Downing
      - Arena Ann Beckham 1825-1903
        - Nathaniel G Beckham
        - Martha Hettie Procter
    - Lucy Jane Ross 1856-1943
      - James J. Ross 1829-1894
        - Joshua Ross
        - Jane
      - Mary Ann Wren 1828-1908
        - James Wren
        - Elizabeth Hightower
  - Ora Ann Elliott 1884-1977
    - James M Elliott 1850-1936
      - James M Elliott 1816-1897
        - Stephen Elliott
        - Elizabeth
      - Margaret Eoff 1824-1890
        - John George Eoff
        - Lucy Shaw
    - Mary C Hinkle 1856-1943
      - John A Hinkle 1822-1897
        - Walter Baltis Hinkle
        - Ann Downing
      - Mary Ann Linn 1828-1893
        - George Linn
        - Temperance Kerwin

Historical Records

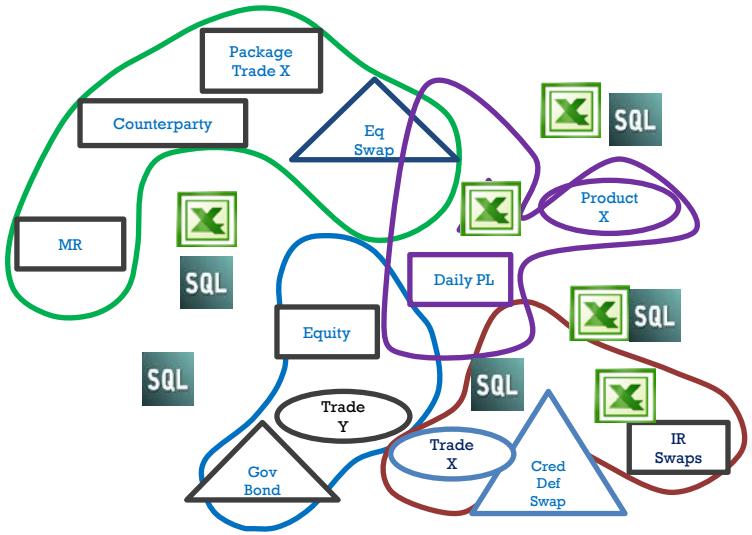
- 1850 United States Federal Census
- 1860 United States Federal Census
- 1870 United States Federal Census
- 1880 United States Federal Census
- Arkansas Census, 1819-70
- Family Data Collection - Births

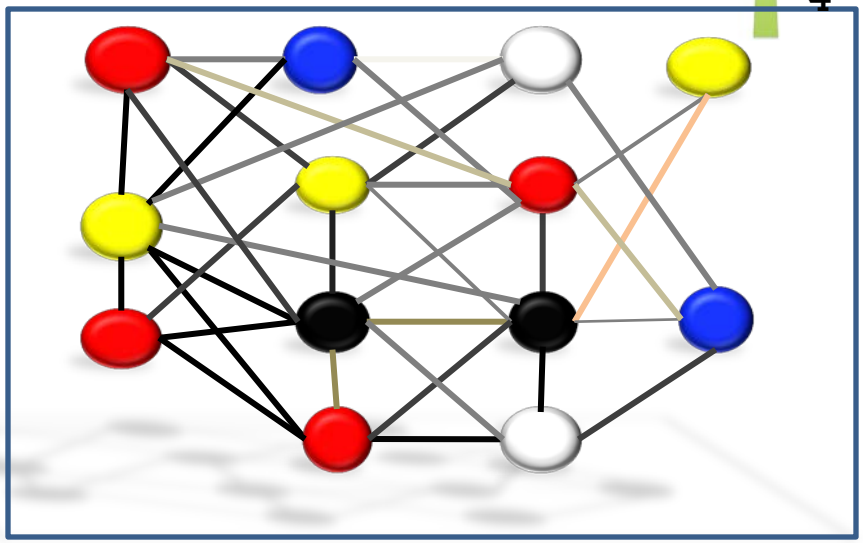
Provenance/sources for tracking family members in the 19th century include early census data (often error prone), military records, passenger & immigration lists, online documents (e.g., county histories, church histories, etc.)

- ***Historical/forensic research requires cross-domain search of a wide variety of resources within a given geo-spatial/temporal context***
- ***Similar capabilities are essential for business intelligence, law enforcement, government applications – all require terminology reconciliation***

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Common challenges for institutions





Current State of Business Data

- Data incongruity and fragmentation often found across silos
- Limited data standards
- Data rationalization problems
- Costly application program logic required to process data into concepts
- Brittle schemas are costly to change
- Rigid and limited taxonomies

Desired State of Business Data

- Data linkage and integration *despite* silos
- Open global *reusable* data standards
- Alignment based on *meaning*
- Highly expressive data schemas with built in *rules* that reflect *concepts*
- Flexible changeable schemas
- Rich multi-level taxonomies

-- courtesy David Newman, Wells Fargo, & Mike Bennett, EDM Council

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Tutorial Outline

- Introduction to Knowledge Representation
- OWL Basics
- Tools & Applications

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Part 1: Introduction to Knowledge Representation

- A little background and a few definitions
- Layers of abstraction & conceptual modeling
- Classifying ontologies
- A little methodology

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Historical Context

- Knowledge Representation
  - Cross-disciplinary field with historical roots in philosophy, linguistics, computer science, and cognitive science
  - Goal is to represent the meaning of knowledge unambiguously, so that it can be understood, shared, and used by computational agents acting on behalf of people to accomplish some task

- Philosophical origins
  - Socrates questioning, Plato's studies of epistemology – the nature of knowledge
  - Aristotle's shift to terminology, development of logic as a precise method for reasoning about knowledge
  - Arguments for the existence of God dating back to Anselm of Canterbury
  - Medieval theories of reference and of mental language, Scholastic logic

Plato and Aristotle at the School of Athens, by Raphael

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Historical Context

- "Brain Cells for Grandmother"
  - Neuroscientists continue to debate today how we store memories
  - One theory, documented as recently as this month in a Scientific American article, suggests that single neurons hold our memories, as concept cells
  - "Each concept – each person or thing in everyday experience – may have a set of corresponding neurons assigned to it"
    - Rodrigo Quian Quiroga, Itzhak Fried and Christof Koch, February 2013 issue, Scientific American
  - "…a relatively few neurons, numbering in the thousands or perhaps even less, constitute a "sparse" representation of an image."
  - " Our brain may use a small number of concept cells to represent many instances of one thing as a unique concept – a sparse and invariant representation … What is important is to grasp the gist of particular situations involving persons and concepts that are relevant to us, rather than remembering an overwhelming myriad of meaningless detail."
  - "The full recollection of a single memory episode requires links between different but associated concepts … If two concepts are related, some of the neurons encoding one concept may also fire to the other one."

Visual perception - Neural coding from the University of Leicester, Department of BioEngineering

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Definitions

- *An ontology* is a specification of a conceptualization. *– Tom Gruber*
- *Knowledge engineering* is the application of logic and ontology to the task of building computable models of some domain for some purpose. *– John Sowa*
- *Artificial Intelligence* can be viewed as the study of intelligent behavior achieved through computational means. Knowledge Representation then is the part of AI that is concerned with how an agent uses what it knows in deciding what to do. *– Brachman and Levesque, KR&R*
- *Knowledge representation* means that knowledge is formalized in a symbolic form, that is, to find a symbolic expression that can be interpreted. *– Klein and Methlie*
- The task of *classifying* all the words of language, or what's the same thing, all the ideas that seek expression, is the most stupendous of logical tasks. Anybody but the most accomplished logician must break down in it utterly; and even for the strongest man, it is the severest possible tax on the logical equipment and faculty. *– Charles Sanders Peirce, letter to editor B. E. Smith of the Century Dictionary*

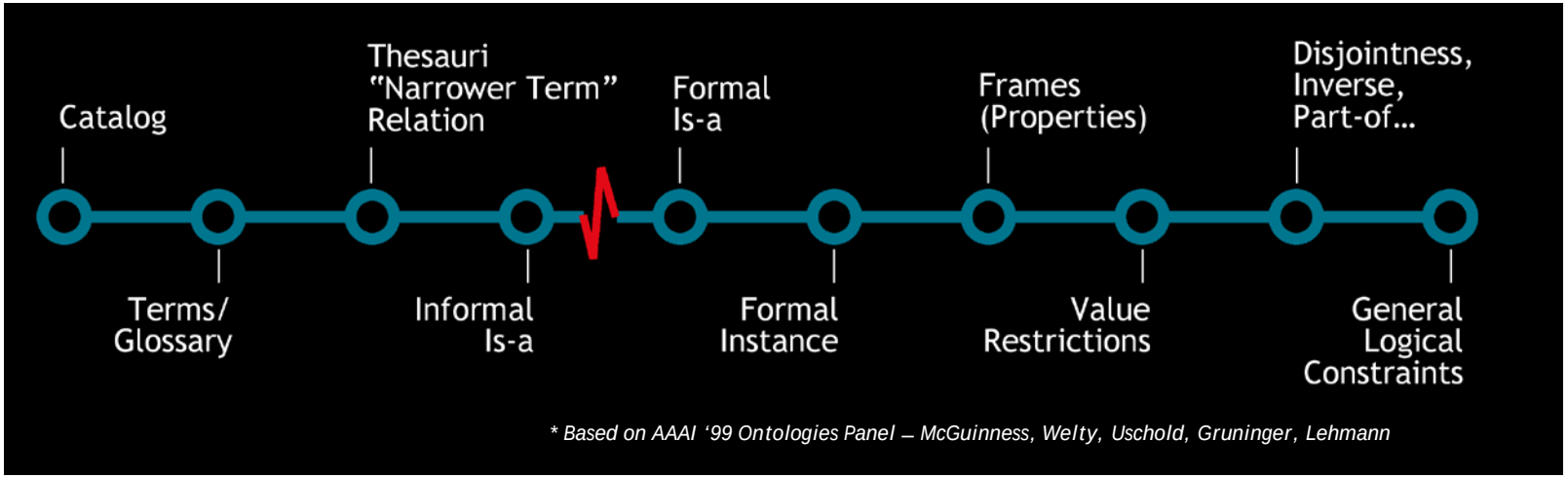
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

What is an ontology?

An ontology specifies a rich description of the

- Terminology, concepts, nomenclature
- Relationships among and between concepts and individuals
- Sentences distinguishing concepts, refining definitions and relationships (constraints, restrictions, regular expressions)

relevant to a particular domain or area of interest.



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Logic and ontological commitment

- Logic can be more difficult to read than English, but is more precise:

(forall ((x FloweringPlant))

(exists ((y Bloom)(z BloomColor))(and (hasPart x y)(hasCharacteristic y z))))

Translation: Every flowering plant has a bloom which is a part of it, and which has a characteristic bloom color.

Language: ISO Common Logic, CLIF syntax

- Logic is a simple language with few basic symbols
- The level of detail depends on the choice of predicates –
  - the predicates represent an *ontology* of the relevant concepts
  - different choices of predicates represent different *ontological commitments*

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Ontology-based technologies

- Ontologies provide a *common vocabulary* for use by independently developed resources, processes, services
- *Agreements* between organizations sharing common services can be specified as formal ontologies, or ontologies with rules, to
  - assist in enforcing explicitly stated policies
  - evaluate usage criteria for services
  - ensure that the *meaning* of relevant concepts is expressed *unambiguously*
- By *composing / mapping* ontologies and *mediating* terminology across participating events, resources and services, independently-developed services can work together to share information and processes consistently, accurately, and completely
- Ontologies also ensure
  - Valid conversations among agents to collect, process, fuse, and exchange information
  - Accurate searching by ensuring context using concept definitions and relations in addition to statistical relevance
  - Policies and rules are consistent with one another to assist in semi-automated policy analysis and enforcement

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

KR language features

- *Vocabulary* –
  - Domain-independent *logical symbols* and reserved terms
  - Domain-dependent *constants*, identifying individuals, properties, or relations in the application domain or universe of discourse
  - *Variables*, whose range is governed by quantifiers
  - *Punctuation* that separates or groups other symbols
- *Syntax* –
  - rules for combining the symbols into well-formed expressions
  - rules may be stated in a linear grammar, graph grammar, or independent abstract syntax
- *Semantics* –
  - a theory of reference that determines how the constants and variables are associated with things in the universe of discourse
  - a theory of truth that distinguishes true statements from false ones
- *Rules of Inference* –
  - rules that determine how one pattern can be inferred from another
  - if the logic is sound, the rules of inference must preserve truth as determined by the semantics

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Classifying logics

Logics vary from classical FOL along a number of dimensions:

- syntax
- the subsets of FOL they implement – for example, propositional logic without quantifiers, Horn-clause, which excludes disjunctions in conclusions such as Prolog, and terminological or definitional logics, containing additional restrictions
- their proof theory:
  - *Intuitionistic logic* and relevance logic rule out certain extraneous information
  - *Non-monotonic logics* allow introduction of default assumptions
  - *Access-limited logic* restricts the number of times a proposition can be used in a proof; *Linear logic* allows a proposition to be used only once
  - *Modal logic* incorporates modal auxiliaries (◊p means p is possibly true; □p means p is necessarily true), temporal logic extends model logic to include always, sometimes
  - *Intensional logics* express concepts such as need, ought, hope, fear, wish, believe, know, expect, and intend
- their model theory, which determines how expressions in the language are evaluated with respect to some model of the world: classical FOL is two-valued; a three-valued logic introduces unknowns; *fuzzy logic* uses the same notation as FOL but with an infinite range of certainty factors (0.0 to 1.0)
- ontology – frameworks may include support for built-in components, such as set theory or time

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Description logic

- A family of logic-based Knowledge Representation formalisms
  - Descendants of semantic networks and frame-based languages such as KL-ONE
  - Describe domain in terms of concepts (classes), roles (relationships), and individuals (instances)
- Distinguished by
  - Formal semantics
    - *Decidable* fragments of FOL
    - Closely related to propositional, modal, and dynamic Logics
  - Provision of inference services
    - *Sound and complete decision procedures* for key problems
    - Implemented systems (*highly optimized*)
- Applications include
  - Configuration – product configurators, consistency checking, constraint propagation, first significant industrial application (*e.g.,* CLASSIC)
  - Question answering and recommendation systems, for suggesting sets of responses or options depending on the nature of the queries
  - Model engineering applications, including those that involve analysis of the ontologies or other kinds of models to determine whether or not they meet certain methodological or other design criteria

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Knowledge bases, databases, and ontology

- An ontology is a conceptual model of some aspect of a particular universe of discourse (or of a domain of discourse)
- Typically, ontologies contain only "rarified" or "special" individuals, *metadata*, representing elemental concepts critical to the domain
- A knowledge base is a persistent repository for
  - Ontology & metadata representing individuals, facts, & rules about how they can be combined or relate to one another
  - Metadata, individuals, facts & rules only – in some applications and frameworks the ontology is separately maintained
- Most inference engines require in-memory deductive databases for efficient reasoning (including commercially available reasoners)
- A knowledge base may be implemented in a physical, external database, such as a relational database, but reasoning is typically done on a subset (partition) of that knowledge base in memory

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Reasoning and truth maintenance

- Reasoning is the mechanism by which the assertions made in an ontology and related knowledge base are evaluated by an inference engine.
  - In classical logic, the validity of a particular conclusion is retained even if new information is asserted in the knowledge base.
  - This may change if some of the preconditions are actually hypothetical assumptions invalidated by the new information.
  - The same idea applies for arbitrary actions – new information can make preconditions invalid.
- Reasoners work by using the rules of inference to look for the "deductive closure" of the information they are given.
  - They take the explicit statements and the rules of inference and apply those rules to the explicit statements until there are no more inferences they can make.
  - When some kind of logical inconsistency is uncovered, then the reasoner must determine, from a given invalid statement, whether or not others are also invalid.
  - The "housekeeping" associated with tracking the threads that support determining which statements are invalidated is called *truth maintenance*.

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Negation

- If all new information is "positive" (monotonic), then all prior conclusions *will*, by definition, remain valid.
- Problems arise if new information negates a prior assumption, causing it to be withdrawn –
  - Conclusive information is not available?
  - The assumption cannot be proven?
  - The assumption is not provable using certain methods?
  - The assumption is not provable given a fixed quantity of time?
- The answers to these questions can result in different approaches to negation and differing interpretations by non-monotonic reasoners.
- Solutions include chronological and "intelligent" backtracking algorithms, heuristics, circumscription algorithms, justification or assumption-based retraction, depending on the reasoner and methods used for truth maintenance.
- Reasoning efficiency is dependent, in part, on the algorithms applied for truth maintenance.

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

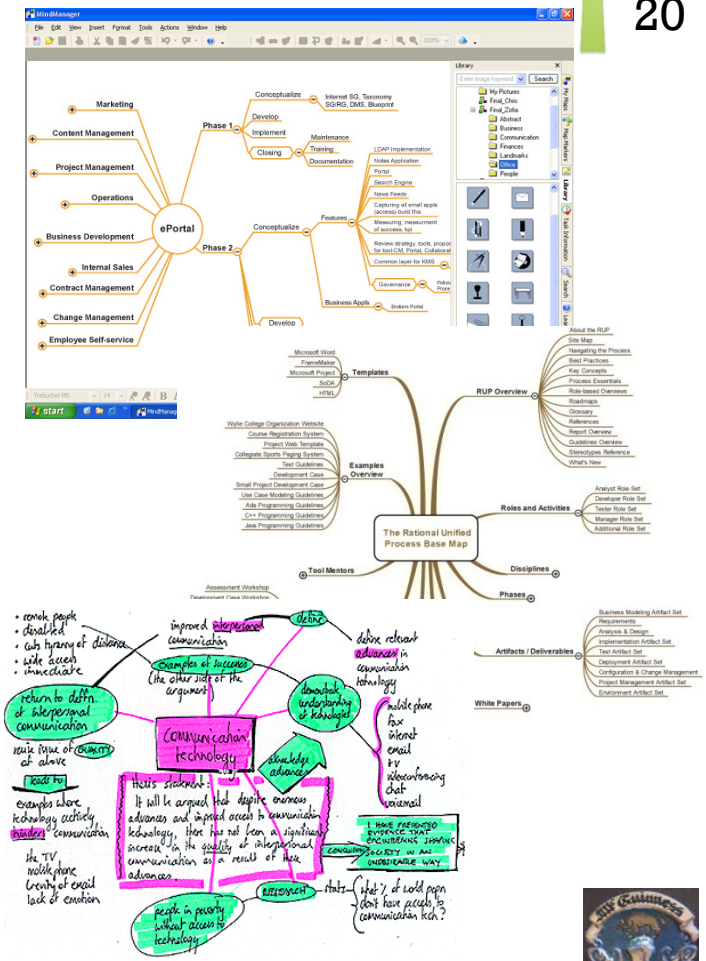
Explanations and proofs

- When a reasoner draws a particular conclusion, many users and applications want to understand *why?*
- Primary motivations include interoperability, reuse, trust, and debugging
- Understanding the *provenance* of the information and results is crucial, especially when web-based information is involved
  - What information sources were used (source)
  - How recently they were updated (currency)
  - How reliable these sources are (authoritativeness)
  - Was the information directly available or derived, and if derived, how (method of reasoning)
- Methods used to explain why a reasoner reached a particular conclusion include explanation generation and proof specification

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Analysis approaches

- Domain analysis – the systematic development of a model of some area of interest for a particular purpose
- The analysis process, including the specific methodology and level of effort required, depends on
  - the context of the work
  - the requirements and use cases relevant to the project
  - the target deliverables
- Approaches to analysis range from high-level mind mapping and brainstorming … to detailed collaboration, dialog, and information modeling to support knowledge sharing
- Common capabilities include
  - "drawing a picture" that includes concepts and relationships between them
  - producing sharable artifacts, that vary depending on the tool – often including web sharable drawings



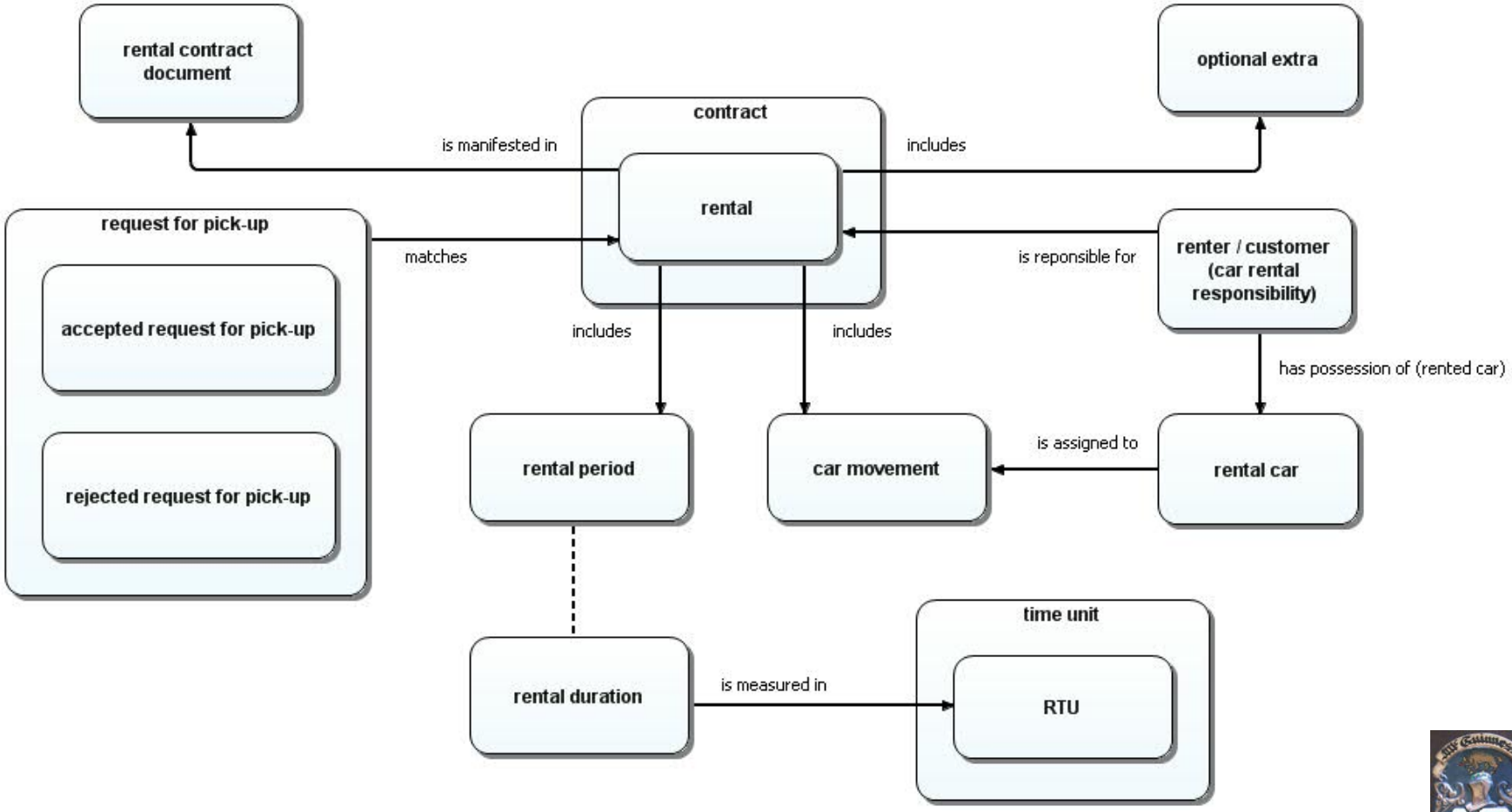
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Abstraction layers

Context	• Identify subject areas	***Business Architecture, Information Architecture & Ontology***
Conceptual	• Define the meaning of things in the domain	
Logical	• Describe the logical representation of concepts, terminology, and their relationships	***Ontology, Entity-Relationship (ER), Business Process Modeling (BPMN), Systems Entgineering (SysML) …***
Physical	• Describe the physical representation of data for repositories and systems	***ER, Relational, XML Schema***
Definition	• Encode the data, rules, and logic for a specific development platform	***XML, source code, scripting languages, stored procedures…***
Instance	• Hold the values of the properties applied to the data in tables in a repository	***Physical KBs, databases, asset repositories…***

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

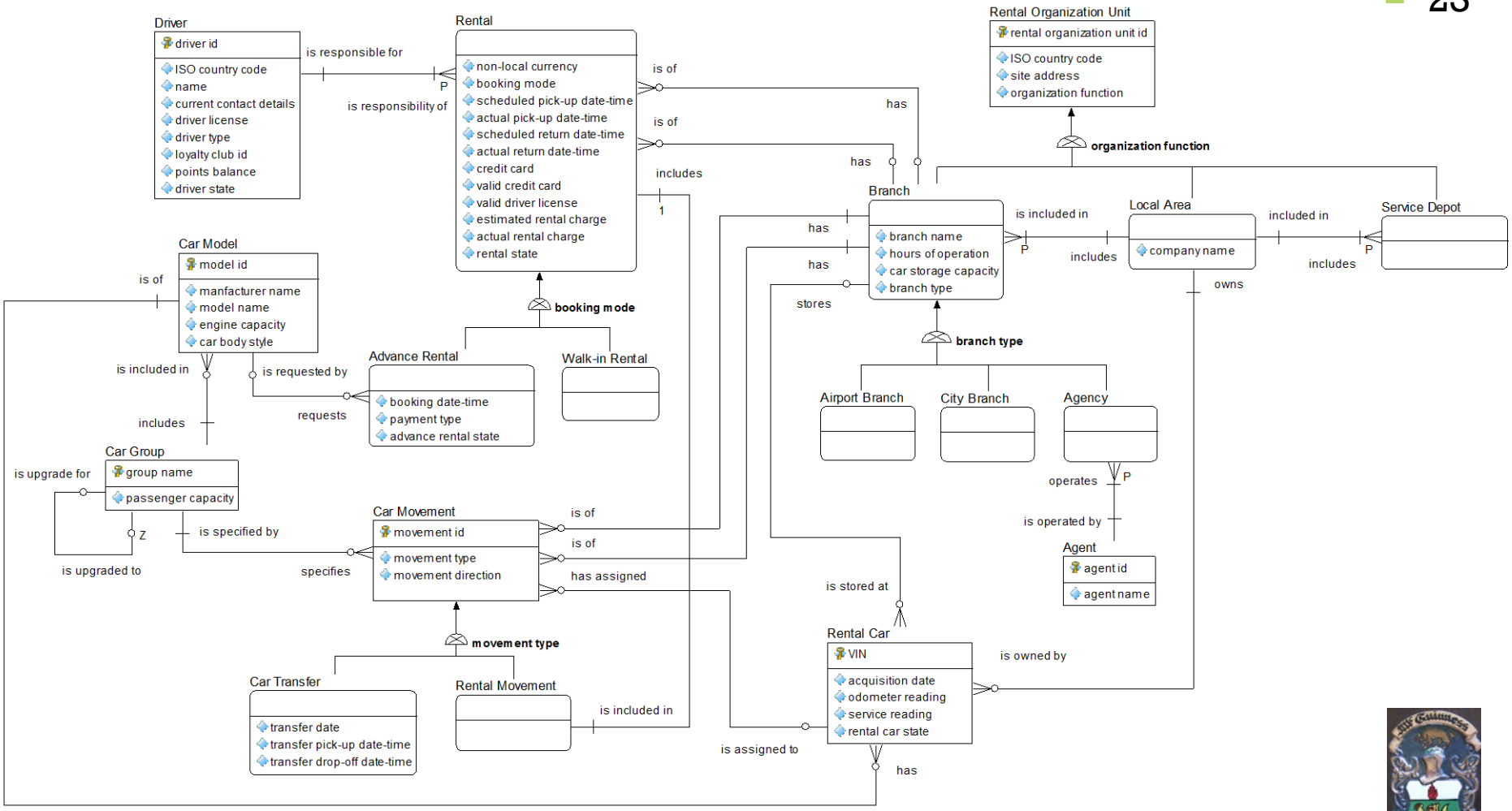
Hypothetical conceptual model "EU-Rent"



Produced using Embarcadero EA/Studio Business Modeler Edition, courtesy Kenn Hussey

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

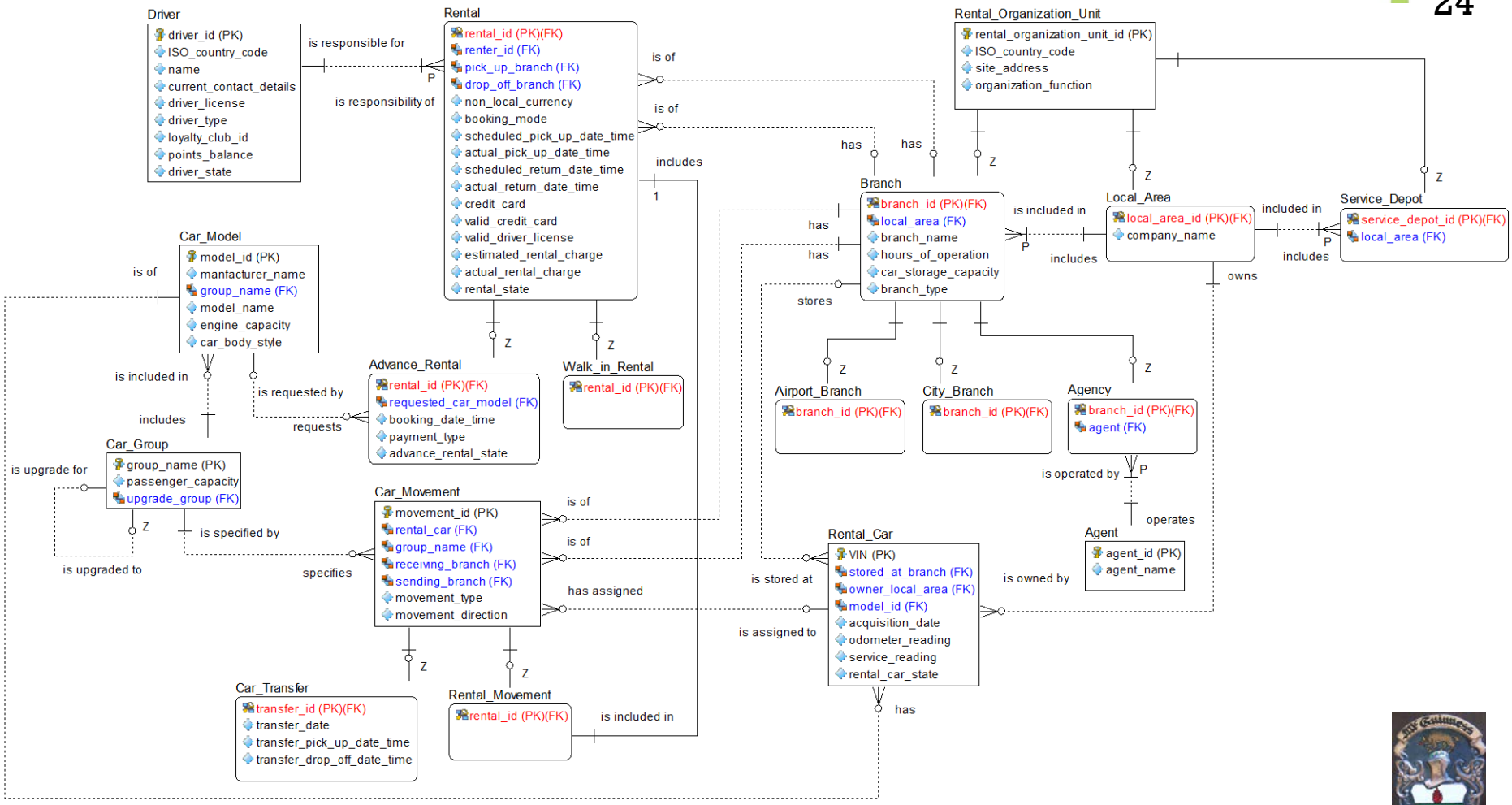
Hypothetical logical model "EU-Rent"



Produced using Embarcadero ER/Studio, courtesy Kenn Hussey

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

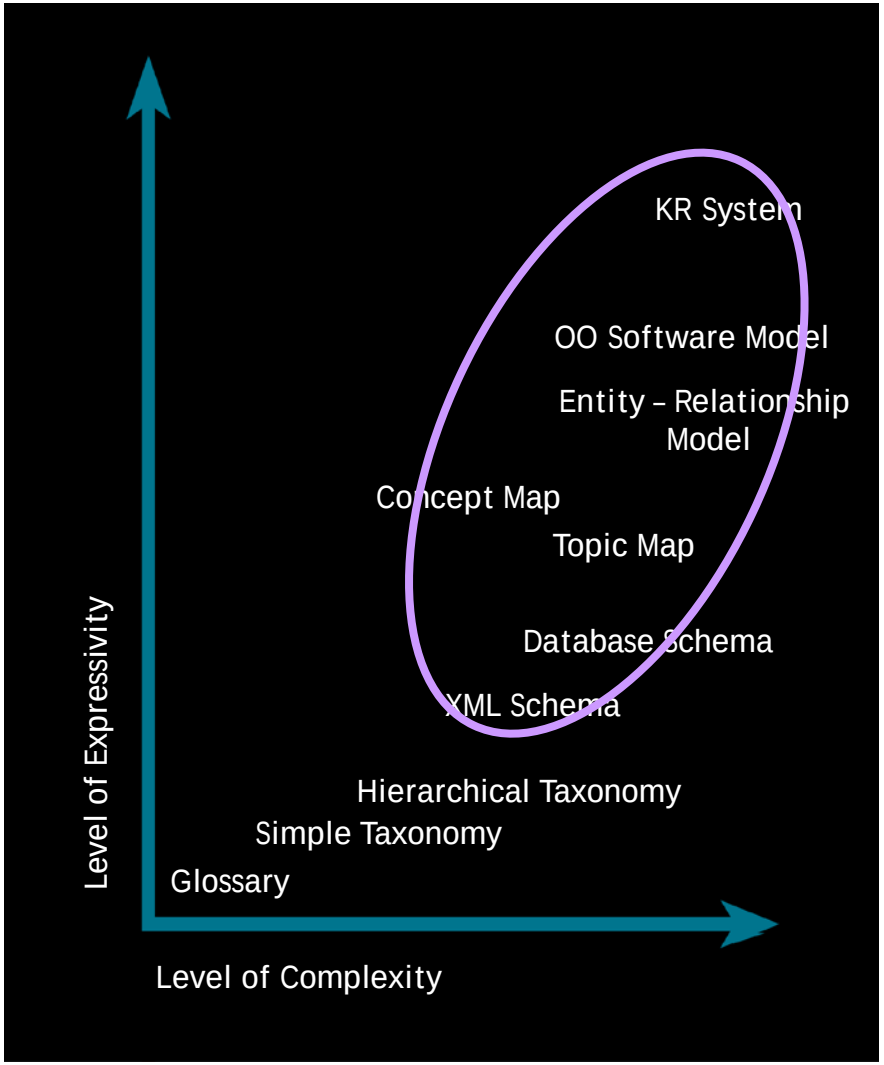
Hypothetical physical model "EU-Rent"



Produced using Embarcadero ER/Studio, courtesy Kenn Hussey

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Classifying ontologies



Classification techniques are as diverse as conceptual models; and generally include understanding

- Level of Expressivity
- Level of Complexity / Structure
- Granularity
- Target Usage, Relevance
- Amount of Automation, Reasoning Requirements
- Prescriptive vs. Descriptive / Reliability / Level of Authoritativeness
- Design Methodology
- Governance
- Vocabulary Management, Metrics

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Considerations

- Intended use of ontologies, including domain requirements (*e.g.*, scientific and engineering apps require formulas, units of measure, computations that may be challenging to represent)
- Intended use of KRSs that implement them, including reasoning requirements, questions to be answered
- For distributed environments, the number and kinds of resources, processes, services requiring ontologies – how distributed, how unique, developed collaboratively or independently, dynamic community participation or static
- What kinds of transformations are required among processes, resources, services to support semantic mediation
- Ontology and KRS alignment / de-confliction / ambiguity resolution requirements
- Ontology and KRS composition requirements, dynamic vs. static composition, in what environment and under what constraints
- Performance, sizing, timing requirements of target environment

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

A little methodology …

- Requirements, domain & use case analysis are critical
  - Develop initial source/reference material
  - Focus on system or application requirements
  - Iterative development starting with a "thread" that covers basic capabilities can ground the work and prioritize decisions
- Need to understand and communicate
  - Architectural trade-offs, cost & technical benefits
  - The nature of the information & kinds of questions that need to be answered drive the architecture, scope and design
- Use discipline from formal domain analysis and use case development to
  - document and explain requirements
  - identify information sources and models (including modularity) needed
  - limit scope creep
- Reuse standards and well-tested, available models whenever possible

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

What to look for

- A controlled vocabulary

  Royal Horticultural Society Color Chart

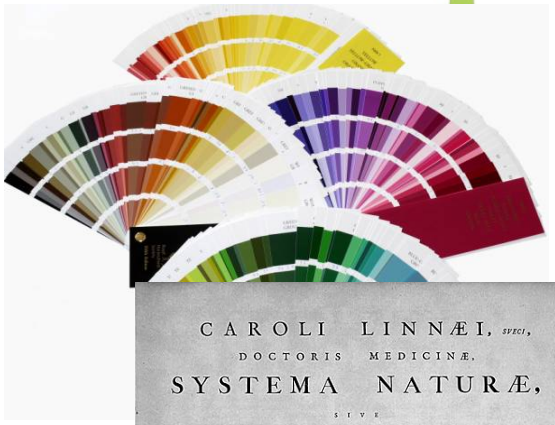
- A hierarchical or taxonomic structure

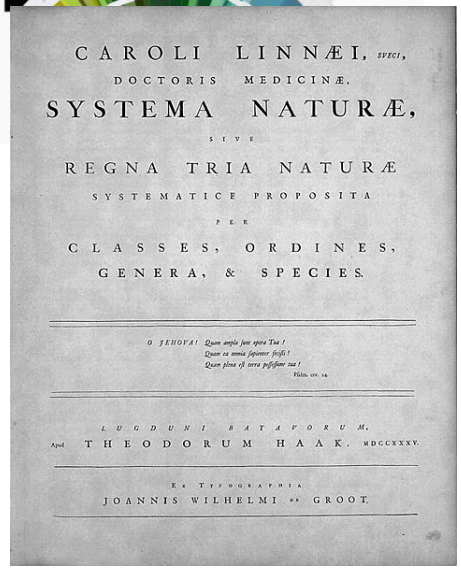
  Linnaean taxonomy, the Plant List – from the Royal Botanic Gardens, Kew, and Missouri Botanical Garden

- Knowledge supporting structured queries (especially for web site, database, information-oriented applications)

  Find deciduous trees native to northern California that are drought tolerant, resistant to oak root fungus, and grow no taller than 20-25 feet

- Requirements specifications
- Reference materials that support domain analysis
- Corporate standards for modeling style as well as content





CAROLI LINNÆI, SVECI,
DOCTORIS MEDICINÆ,
SYSTEMA NATURÆ,
SIVE
REGNA TRIA NATURÆ
SYSTEMATICE PROPOSITA
PER
CLASSES, ORDINES,
GENERA, & SPECIES.

O JEHOVA! Quam ampla sunt opera Tua!
Quam ea omnia sapienter fecisti!
Quam plena est terra possessione tua!
Psalm. civ. 24.

LUGDUNI BATAVORUM,
Apud THEODORUM HAAK. MDCCXXXV.

EX TYPOGRAPHIA
JOANNIS WILHELMI DE GROOT.

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Using use cases to gather requirements

- A good summary for every use case should include:
  - A description of the basic business requirement / need the use case is intended to support
  - Primary goals
  - Scope – identify any known boundaries as a starting point
  - Pre-conditions and post conditions – any assumptions you know about the state of the "system"/world before and after
  - Actors and interfaces – identify primary actors, information sources, interfaces to existing or new systems
  - Triggers – what kicks off the use case, any particular series of events, situation, activity, etc., and any that affect the flow
  - Performance requirements – including any sizing or timing constraints, "ilities" , etc.

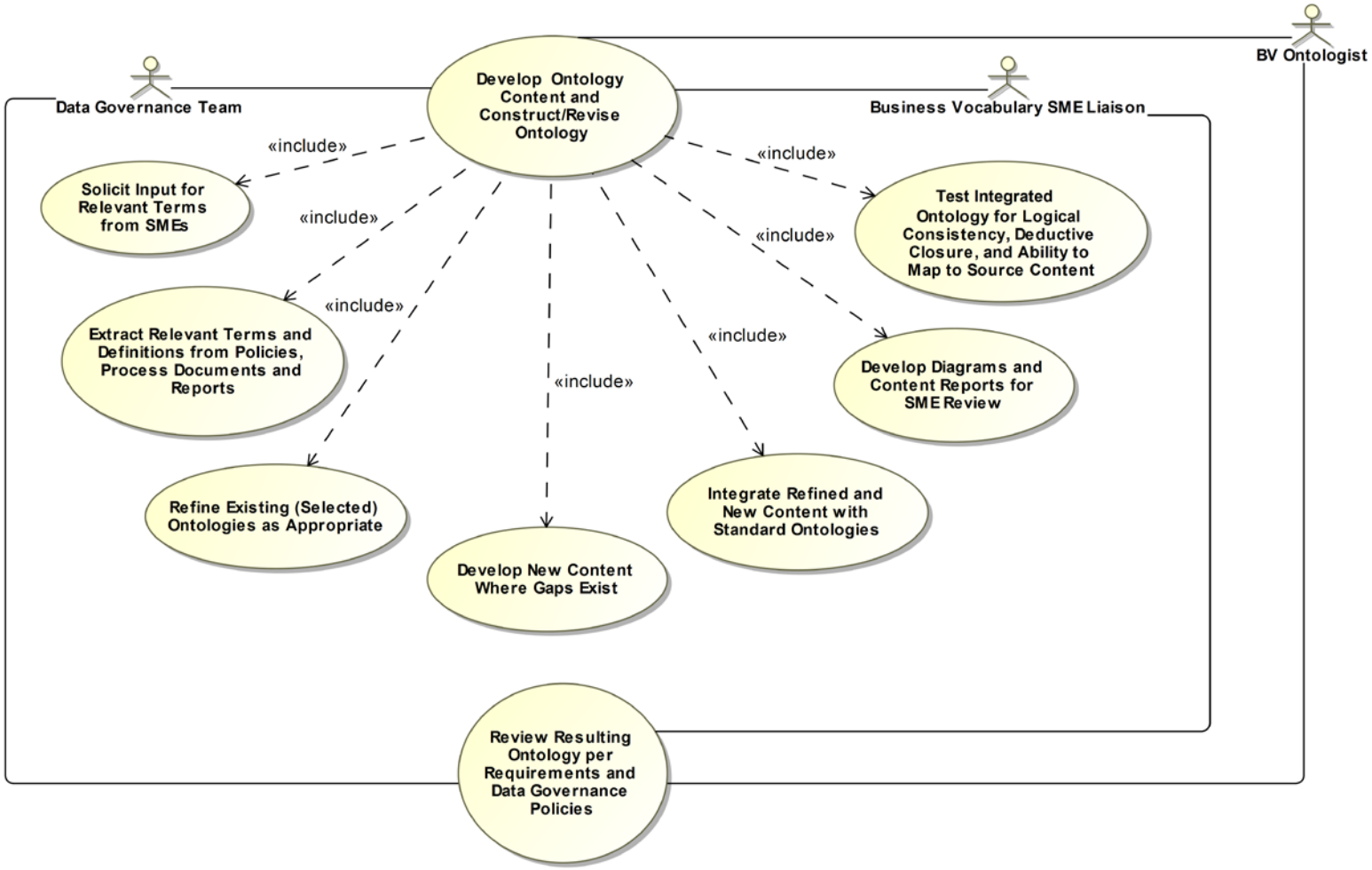
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Use case content

- Outline the major process steps for both the "normal", or primary scenario, and alternative flows, such as if things don't go well
- Use case and activity diagrams – typically done in UML, but could be Visio, PowerPoint, or whatever tools your team is comfortable with
- Usage scenarios – you should have at least two narrative "stories" that describe how one of the main actors would experience the use case, with the intent of identifying additional requirements
- Competency questions – identify as many of the questions you want the knowledge base / application to answer as possible
- Resources – describe any known contributing resources and repositories, other external systems that participate in the use case to the degree possible

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Example partial use case diagram for content development process



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Start with canonical definitions

- General concepts as well as domain-specific knowledge
- Basic starting point – cross-domain definitions
  - Namespace definitions (as applicable), metadata, naming conventions, governance policies
  - Commonly used structures & vocabularies, such as
    - messaging interface standards
    - common API structures
    - corporate standard business glossaries and vocabularies
    - corporate or departmental business architecture and process models
    - domain-independent standards for accounting, currencies, reporting
  - Common metadata for model, schema, and ontology management (*e.g.*, Dublin Core, for documents & models, ISO 1087 for synonyms & similar relations, MIME media types for images & multimedia, SKOS for annotations, etc.)
- Domain vocabularies must be prioritized, selected based on business requirements, clear ROI

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Basic domain analysis

- Identify the main concepts in the domain – classes / terms
  - For a trading system – trade, trader, trading platform, financial instrument, contract, counterparty, price, market valuation …
  - For a nursery example – landscape, light requirements, dimensions, plants, planting material, hardscape material, containers …
- Identify any structural relationships between the concepts – both hierarchical and lateral relationships
  - Specialization / generalization (parent / child) – always true is-a hierarchies, no "cheating" on this
  - isPartOf / hasPart, isMemberOf / hasMember, other mereologic relations
  - Other functional, structural, behavioral relationships
  - Key attributes – dates/times, identifiers, etc.

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Capturing definitions

- Layout a high-level architecture for key ontologies and ontology elements
- Identify the relationships among elements – roles, domain, interface, process, utility
- Define an approach for gathering content from subject matter experts, possibly based on IDEF5 (Integrated Definition Methods) Ontology Capture Method Analysis, that includes
  - Understanding and documenting source materials
  - An interview template
  - Traceability back to your use cases
- For each ontology element
  - Describe its domain and scope, how it will be used
  - Identify example questions and anticipated/sample answers for the application(s) it will support
  - Identify key stakeholders, ownership, maintenance, resources for instance knowledge
  - Describe anticipated reuse/evolution path
  - Identify critical standards, resources that it must interoperate with, dependencies
- Resources
  - http://www.idef.com/IDEF5/html
  - http://www.kbsi.com/technology/methods/sbont.htm

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Terminology analysis

- ISO 704, Principles & Methods for Terminology Work, provides a methodology for describing concepts & terms
  - Uses ISO 1087 for terminology
  - Uses ISO 860 for terminology "harmonization" (alignment) methods
  - Basis for typical methods used for taxonomy development today
- Describes how to flesh out definitions – Aristotelian genus/differentia structure
  - For classes – "a <parent class> that …", including text that provides content you'll later express as restrictions, other refinement
  - For properties – "a <parent property> relation [between <domain> and <range>] that …"
  - SBVR style formal definitions (Semantics of Business Vocabularies and Rules – see http://www.omg.org/spec/SBVR/
- Recommendations strategies for relating terms to one another using standard vocabulary
- ISO 1087 – great resource for language to describe kinds of relationships, acronyms & other designations, preferred vs. deprecated terms, etc.
- ISO 860 augments this with recommendations for vocabulary comparison

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Successful strategies …

- Successful ontology/vocabulary development model reflects small development team with broader user Community of Interest / stakeholders, especially for reusable content models
  - Provide readable documentation, even for small communities
  - State maintenance policies clearly
  - Identify versions
  - Publish models for ease of accessibility by COI members
- Where different stakeholder communities have unique requirements, explicitly specified models can be mapped to one another, translation services developed
  - Common terminology services (CTS2) standard from the OMG for healthcare

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Common applications

- Canonical information definitions – shared representation & understanding facilitates service interoperability, reuse
- Asset/artifact repository search & retrieval
- Service registration, description, discovery & management
- Augmented decision support and business rule applications
- Better web site / application architecture
  - Smarter search capabilities (*e.g.*, faceted search, pull-focused applications)
  - Customer experience & cross-sell / upsell (push) – recommendation systems
- Richer interoperability among trading partners – extensions to messaging agreements
- Automated verification

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Separation of concerns / modularity

- Critical dimensions aid in determining module boundaries –
  - separate business-related content from technical detail
  - aspects of the business content, such as marketing and branding, from others describing products or manufacturing processes
  - separate disciplines into independent modules
- Other considerations include separation based on
  - back-end store / source repositories
  - application boundaries, system interfaces
  - distributed resources
  - the need to reason over some parts of the knowledge base but not others to answer sets of critical questions
  - performance requirements, for reasoning, query answering, etc.
  - asserted vs. inferred content

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Naming and versioning

- Naming conventions and versioning policies are critical for –every– organization
  - Namespace definitions for ontologies are critical, along with policies for their management that are well understood
  - http://<authority>/<subdomain>/<topic>/<date, in YYYYMMDD form>/filename.extension is common practice in some organizations, with content negotiation point to the latest version
  - Levels of hierarchy may be added for large organizations or modularized ontologies
  - Namespace prefixes (abbreviations) for individual modules, especially where there are multiple modules, can be important
  - If you post something at a URL, the idea is that it should be permanent, so namespace design and governance is essential

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Best practices in namespace development

- *Availability* – people should be able to retrieve a description about the resource identified by the URI from the network (albeit internally)
- *Understandability* – there should be no confusion between identifiers for networked documents and identifiers for other resources
  - URIs should be unambiguous
  - URIs are meant to identify only one of them, so one URI can't stand for both a networked document and a real-world object
  - Separation of concerns in modeling "subjects" or "topics" and the objects in the real world they characterize is critical, and has serious implications for designing and reasoning about resources
- *Simplicity* – short, mnemonic URIs will not break as easily when shared for collaborative purposes, and are typically easier to remember
- *Persistence* – once a URI has been established to identify a particular resource, it should be stable and persist as long as possible
  - Exclude references to implementation strategies, as technologies change over time (e.g., do not use '.php' or '.asp' as part of the URI scheme), and organization lifetime may be significantly shorter than that of the resource
- *Manageability* – given that URIs are intended to persist, administration issues should be limited to the degree possible
  - Some strategies include inserting the current year or date in the path so that URI schemes can evolve over time without breaking older URIs
  - Create an internal organization responsible for issuing and managing URIs, and corresponding namespace prefixes

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Model element naming

- For naming entities in an ontology – there are some rules of thumb that vary by community of practice
  - data modelers often use underscores at word boundaries, spaces in names – which semantic web tools may not handle well (ok in labels)
  - some name properties <domain><predicate><range>, others <predicate><range>, & others <predicate>, but not necessarily consistently
  - semantic web practitioners typically use camel case (upper camel case for classes, lower camel case for properties)
  - using verbs to the degree possible for property naming, without incorporating the domain/range is preferable
  - for very large ontologies, some use unique identifiers to name concepts and properties, with human readable names in labels only – depends on tooling that helps people interpret the content
- The key is to establish & consistently use guidelines tailored to your organization

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Metadata for ontologies and elements

- Metadata should be standardized at both the model level and the element level for *every* model (not just ontologies)
  - Model level metadata can reuse properties from the Dublin Core Metadata Terms, from the Simple Knowledge Organization System (SKOS), ISO 11179 Metadata Registry standard with ISO 1087 Terminology support, W3C Prov-O vocabulary for provenance and others
  - Most annotations should be optional at the element level, but a minimal set, including names, labels, and formal, text definitions, is important for reusability & collaboration
  - Consistent use of the same annotations (properties, tags) improves readability, facilitates automated documentation generation, and enables better search over ontology repositories
  - Model level metadata may reuse organization-specific taxonomies to enable better search through RDFa tagging, for example
  - Latest version of an OMG architecture board recommended vocabulary for specification metadata, and related annotations from the Financial Industry Business Ontology (FIBO – OMG) are available at
    - http://www.omg.org/techprocess/ab/20130801/SpecificationMetadata.rdf
    - http://www.omg.org/spec/EDMC-FIBO/FND/1.0/Final/ – see AnnotationVocabulary.rdf

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Change management

- Change management and traceability is not well defined at the element level for ontologies, still a research topic
  - Current approach to element-level versioning to support reasoning about change is to track two parallel streams: one for additions to the ontology, one for retractions; often managed as two separate files
  - UML/MOF versioning suggests linking to a workspace; use of a default workspace would get the latest version, and tools such as MagicDraw© & repositories such as Adaptive can provide an indication of the differences
  - UML-based versioning does not support reasoning about the differences so that users can determine the impact of applying the changes
    - addition or deletion of axioms can change downstream reasoning results, especially where there are complex dependencies
  - Mechanisms that preserve the versioning detail in artifacts that are generated from such models, such as RDF/XML serialized OWL, and are compatible with the above approaches, is an ongoing topic of discussion
    - Consider the use cases/justification for whatever level of versioning detail is needed on a project by project basis
    - Balance with usability/performance

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Part 2: OWL Basics

- A quick intro to RDF and RDF Schema
- Basic constructs: descriptions & classes
- Class expressions
- Properties & restricting property values
- Individuals & data ranges
- Miscellaneous class axioms
- Additional features of OWL 2
- A few rules of thumb
  - Depth vs. breadth, naming, synonymous terms
  - Class vs. property value, class vs. individual
  - Limiting scope

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Semantic Web

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

-- Tim Berners-Lee

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Guiding Principles

- Historically, knowledge representation and reasoning systems have operated under *closed world* assumptions
- Uncertainty is magnified under *open-world*, "wild, wild web" conditions, making reasoning much more difficult
- Semantic web languages are designed to support less certainty, to provide "better" search results, informed answers to questions, not absolutes
- Because they are based on XML, such languages can assist businesses in leveraging existing investment in mark-up, content, and data
  - To augment business intelligence/analysis and knowledge mining
  - To support knowledge sharing and collaboration, augment enterprise information integration
  - Enrich web services and other applications
  - Support policy-based applications and ensure compliance with policy

at a lower cost with higher potential ROI than traditional computing methods

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

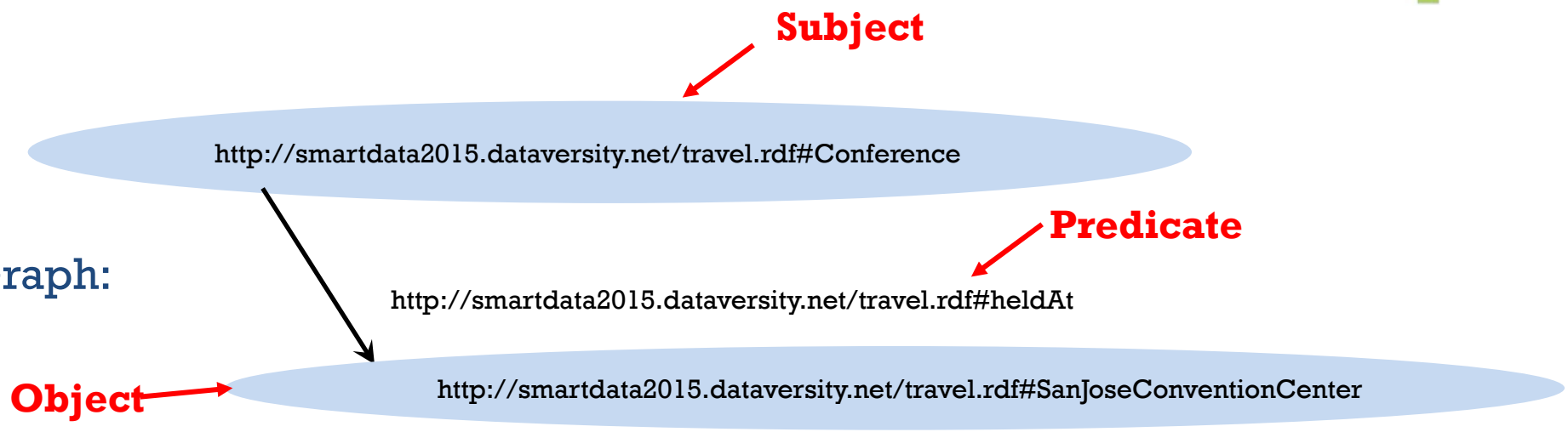
Resource Description Framework (RDF)

- Describes relationships
- Uses URIs used for naming
- Language has
  - graph based model
  - RDF/XML serialization (exchange syntax)
- Specification, W3C presentations, tools are available at
  - Semantic Web: http://www.w3.org/standards/semanticweb/
  - Linked Data: http://www.w3.org/standards/semanticweb/data
  - RDF: http://www.w3.org/standards/techs/rdf#w3c_all
- Recent revisions to RDF include:
  - RDF 1.1 – cleans up a number of issues in the earlier specification, adds support for RDF "sources", such as SPARQL endpoints, and "datasets"
  - Serialization standards – RDF 1.1 Turtle, RDF 1.1 N-Quads, in addition to RDF/XML

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

RDF Notation Options

Graph:



XML/RDF:

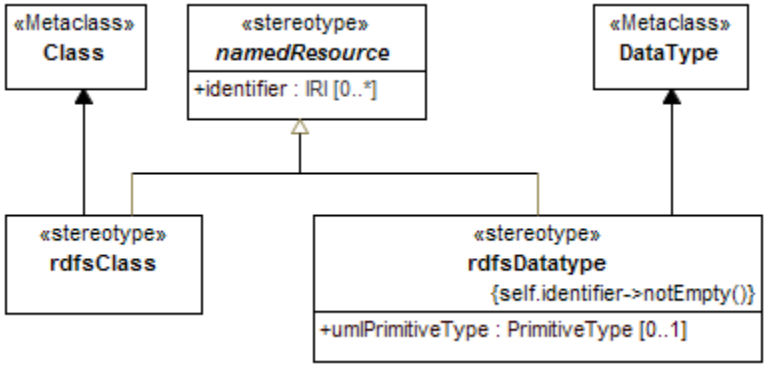
```
<rdf:Description rdf:ID="Conference"/>
  <smartdata2015:heldAt rdf:resource="#SanJoseConventionCenter"/>
</rdf:Description>
```

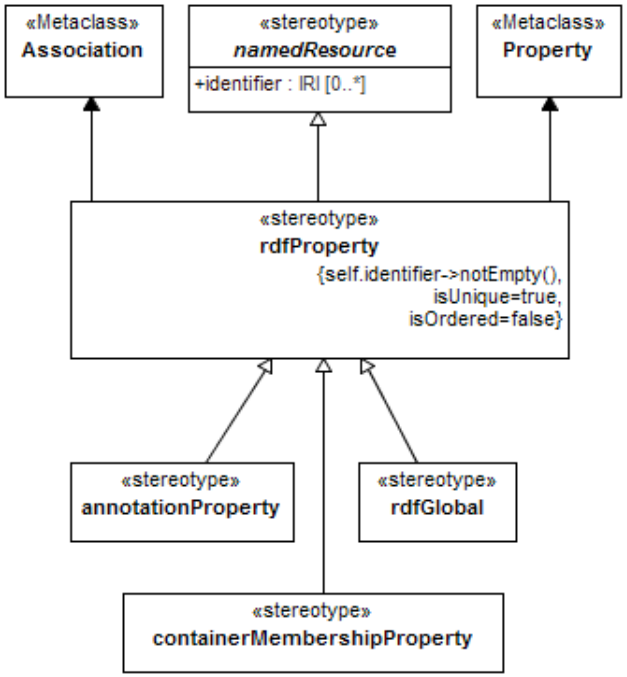
N3: smartdata2015:Conference smartdata2015:heldAt smartdata2015:SanJoseConventionCenter.

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

RDF Schema (RDFS)

- An RDF vocabulary that provides for identifying:
  - classes,
  - subsumption (inheritance) relations for classes,
  - subsumption (inheritance) relations for properties,
  - domain and range for properties

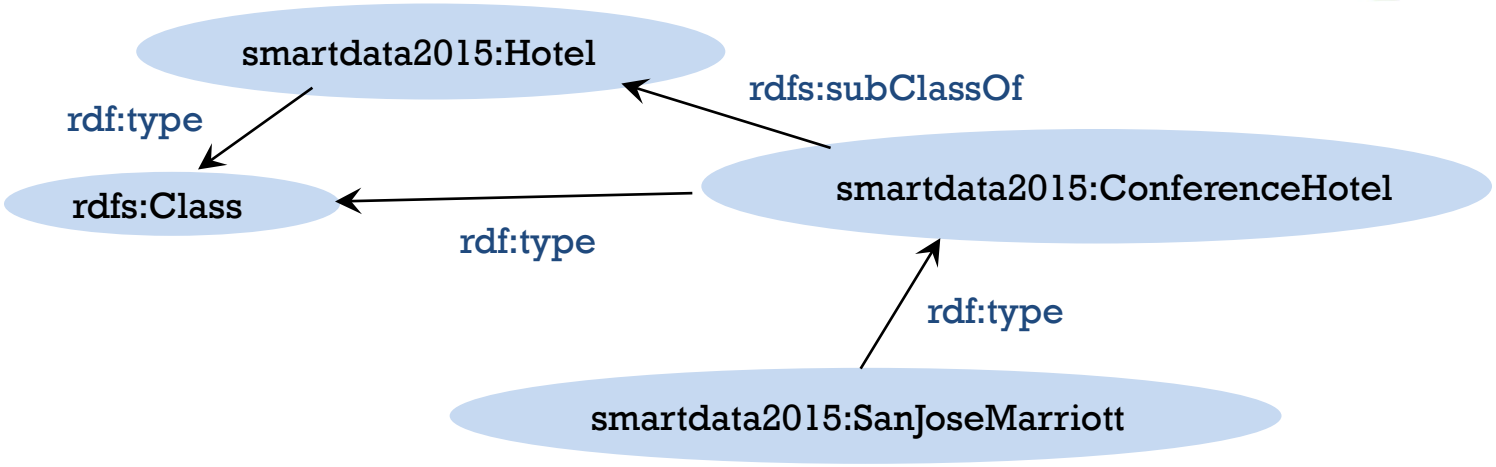




Copyright © 2015 Thematix Partners LLC & McGuinness Associates

RDF Schema (RDFS)

Graph:



XML/RDF:

```
<rdf:Description rdf:ID="Hotel">
   <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
 </rdf:Description>

<rdfs:Class rdf:ID="ConferenceHotel">
   <rdfs:subClassOf rdf:resource="#Hotel"/>
 </rdfs:Class>

<smartdata2015:ConferenceHotel rdf:ID="SanJoseMarriott"/>
```

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

The Web Ontology Language (OWL)

- Two languages emerged in parallel to address semantic web requirements
  - DAML-ONT, supported by the DARPA/DAML program
  - OIL (Ontology Inference Layer) developed by EU & US researchers
- Merged DAML+OIL was submitted to the W3C 2002, formed the basis for the WebOnt Working Group
- OWL extends RDF Schema
  - Has an RDFS based syntax and reuses some RDF vocabulary (*e.g.*, subClassOf, domain, range)
  - Adds rich primitives and redefines others (transitivity, inverse, cardinality constraints, complex class definitions)
- Describes the structure of a domain in terms of classes and properties
- Uses RDFS for class/property membership assertions (ground facts), XML Schema Datatypes
- OWL specifications became W3C recommendations in February 2004
- OWL 2 specifications was adopted in October 2009, with minor revisions in December 2012

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

General nature of descriptions

a WINE

a LIQUID
a POTABLE

grape: chardonnay, ... [>= 1]
sugar-content: dry, sweet, off-dry
color: red, white, rose
price: a PRICE
winery: a WINERY

grape dictates color (modulo skin)
harvest time and sugar are related

General Categories

Structured Components

Interconnections Between Parts

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

General nature of descriptions

Class

a WINE

Superclass

a LIQUID
a POTABLE

General Categories

Number / Card Restrictions

grape: chardonnay, ... [>= 1]
sugar-content: dry, sweet, off-dry
color: red, white, rose
price: a PRICE
winery: a WINERY

Structured Components

Roles / Properties

Value Restrictions

grape dictates color (modulo skin)
harvest time and sugar are related

Interconnections Between Parts

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Describing classes

- A class is a concept in the domain
  - Vintage – *a wine made from grapes grown in a specified year*
  - A set of characteristics (*flavor, body, color, sugar*…)
- A class is a collection of elements with similar properties
  - White wine – *wine made from white grapes*
  - White table wine – *wine made from white grapes that are not appellations or regional* (not "quality wine" in the EU)
- A class expression defines necessary conditions for membership (*specific varietal, field, micro-climate, date picked, date bottled*)
- Instances of classes
  - Silver Oak 1996 Alexander Valley Cabernet Sauvignon
  - Andrew Murray Vineyards Espérance, Central Coast 2012
  - Robert M. Parker, Jr,'s Wine Advocate review, dated February 28, 2002
  - Los Olivos, Santa Barbara County, California, USA

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Class inheritance

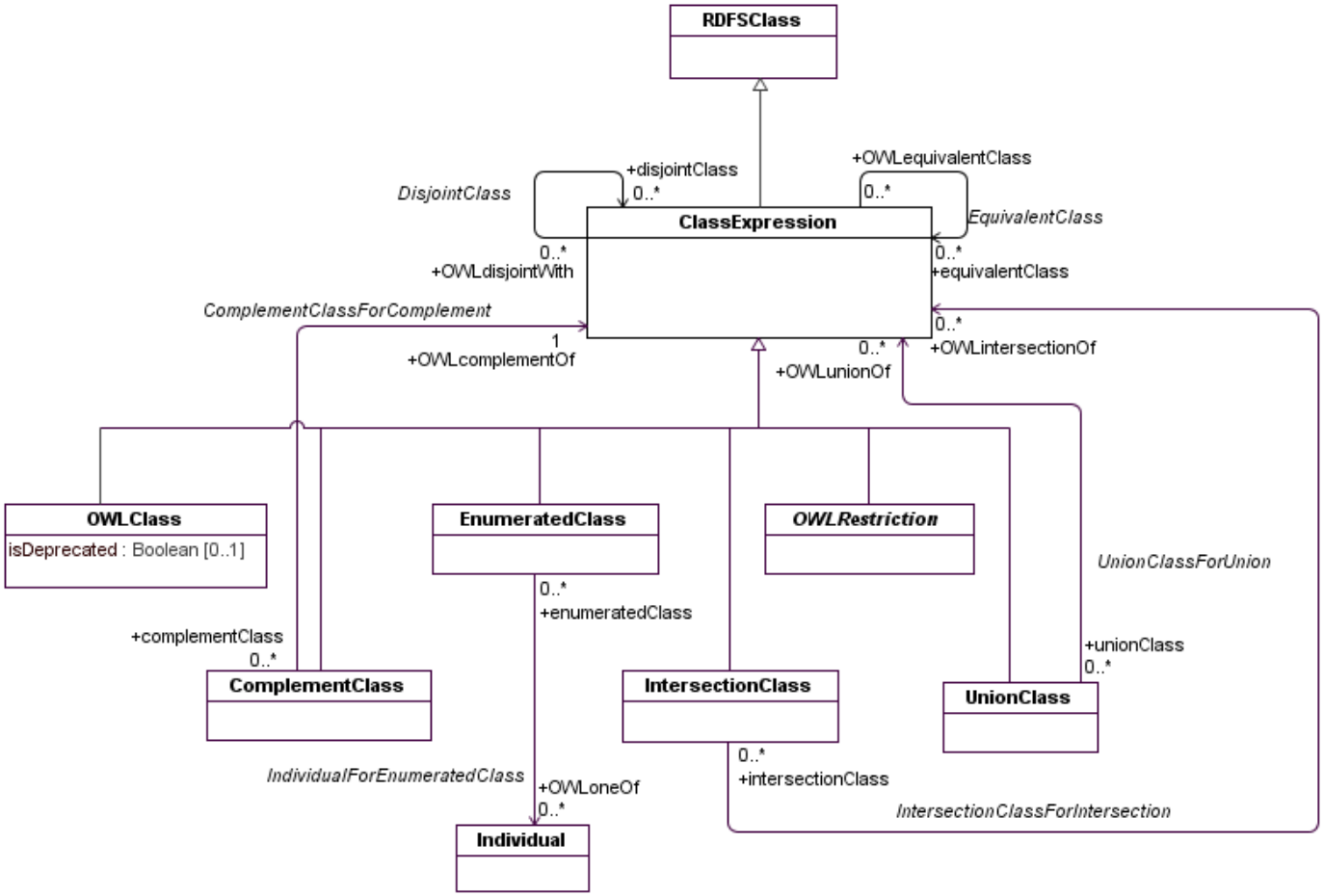
- Classes are organized into subclass-superclass (or generalization-specialization) hierarchies
- True subclass relationships are the basis of a formal *is-a* hierarchy
  - Classes are "is-a" related if an instance of the subclass is an instance of the superclass
  - is-a hierarchies are essential for classification
  - Violation of true is-a hierarchical relationships can have unintended consequences & and cause errors in reasoning
- Class expressions should be viewed as sets, and subclasses as a subset of the superclass, in contrast with a collection of attributes as classes are sometimes specified in object oriented programming
- Examples
  - FloweringPlantType is a subclass of PlantType

    Every flowering plant is a plant; every instance of a flowering plant (e.g., Azalea indica 'Alaska' (Rutherfordiana hybrid) is an instance of a flowering plant
  - Azalea is a subclass of FloweringPlantType, Rhododendron
  - Monrovia is a company that breeds, grows, and sells flowering plants

** courtesy Monrovia web site*

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

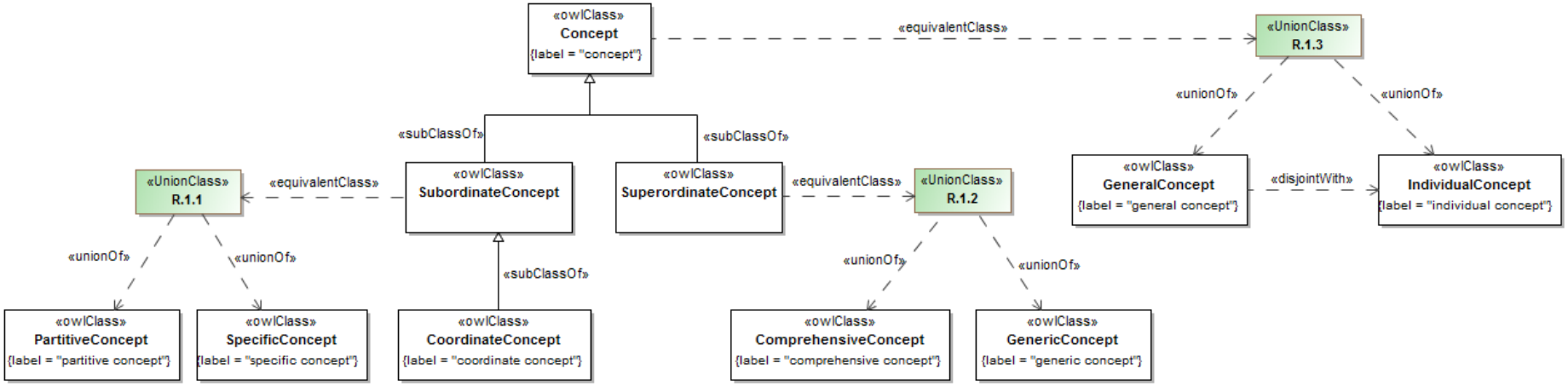
Class expressions



- 6 primary kinds of class expressions in OWL
- 5 anonymous
- Definitions can be nested using set-theoretic constructs

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Example class hierarchy with other expressions



- Example from ISO 1087 combining subclass-superclass and union expressions
- Provides a more accurate representation of the relationships defined in the ISO standard than a strict is-a hierarchy would

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Example class hierarchy with other expressions

File Edit View Reasoner Tools Refactor Window Help

Concepts (http://www.omg.org/spec/VTW/20140501/Concepts/)

Search for entity

Active Ontology | Entities | Classes | Object Properties | Data Properties | Individuals | OWLViz | DL Query | OntoGraf

Class hierarchy | Class hierarchy (inferred)

Class hierarchy (inferred): Concept

- owl:Thing
  - Characteristic
  - Concept
    - GeneralConcept
    - IndividualConcept
    - SubordinateConcept
      - CoordinateConcept
      - PartitiveConcept
      - SpecificConcept
    - SuperordinateConcept
      - ComprehensiveConcept
      - GenericConcept
  - ConceptField
  - ConceptSystem
  - Context
  - Definition
  - Designation
  - Diagram
  - Extension

Individuals by type | Annotation property hierarchy | Datatypes

Object property hierarchy | Data property hierarchy

Object property hierarchy:

- owl:topObjectProperty

Class Annotations | Class Usage

Annotations: Concept

Annotations

rdfs:label
concept

skos:definition [type: xsd:string]
unit of knowledge created by a unique combination of characteristics

skos:note [type: xsd:string]
Concepts are not necessarily bound to particular languages. They are, however, influenced by the social or cultural background which often leads to different categorizations.

sm:directSource [type: xsd:string]
ISO 1087-1, paragraph 3.2.1

Description: Concept

Equivalent To

GeneralConcept
or IndividualConcept

SubClass Of

hasCharacteristic only Characteristic
hasDefinition only Definition
hasDesignation only Designation
hasExtension only (extends exactly 1 Concept)
hasIntension only Intension

SubClass Of (Anonymous Ancestor)

Reasoner active | Show Inferences

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Modeling considerations for class hierarchy development

- Practitioners tend to start
  - Top-down - define the most general concepts first and then specialize them
  - Bottom-up - define the most specific concepts and then organize them into more general classes
  - Combination (typical – breadth at the top level and depth along a few branches to test design)
- Class inheritance is *transitive*
  - A is a subclass of B (white wine, dessert wine are subclasses of wine)
  - B is a subclass of C (sauvignon blanc is a subclass of white wine, late harvest wine is a subclass of dessert wine)
  - therefore A is a subclass of C (late harvest sauvignon blanc is a subclass of white wine, dessert wine, & wine)
  - Start with a more formal is-a approach, tease out whether other expressions are more appropriate based on use cases, formal definitions when available, etc.

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Properties

- Properties describe characteristics, features, or attributes of the members of a class

  Every flowering plant has a bloom color, color pattern, flower form, petal form, etc.

- Classes of properties
  - *intrinsic*: properties of the plant itself such as the optimal sunlight, soil conditions, expected height, and so forth
  - *extrinsic*: properties imposed externally such as the grower and price
  - whole-part relations
  - geospatial, mereonomic relations: what microclimates are this plant well suited to, where in a particular historic garden will you find it

- Data and object properties
  - simple attributes typically have primitive data values (e.g., strings, numbers)
  - complex properties refer to other entities (*e.g.*, an individual grower, such as Monrovia, or organization such as the American Orchid Society)
  - OWL reasoning requires strict separation of data and object properties

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Domain & range properties

- In OWL and many other KR languages, relations (properties) are strictly binary
- The domain & range represent the source & target arguments, respectively, for the property
- Domain – the class (or classes) that may have the property – *Wine is the domain of the property hasWineColor*
- Range – the class (or classes) defining valid property values – everything that fulfills the *hasWineColor* property is an instance of the enumerated class {*red, white, rose*}
- Some KR languages that inherently support n-ary relations, such as CL, do not make this distinction
  - More flexible, intuitively more like mathematics, where functions have ranges (or return types) but not all relations are functions
  - Requires additional relations to specify argument order, which can be critical for ontology alignment

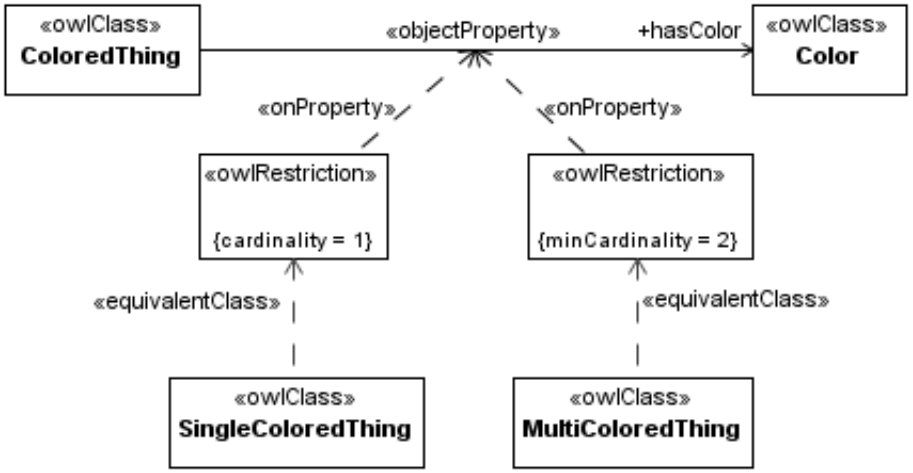
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Class expressions that restrict property values

- Number restrictions describe or limit the number of possible values a particular property can have
  - A language must be associated with at least one English name and at least one French name
  - A language may be associated with zero or more Indigenous names
- Cardinality – similar meaning to classical set theory, measures the number of elements in the set (restriction class)
  - Cardinality – cardinality *N* means the class defined by the property restriction must have exactly *N* values (individual or literal values)
  - Minimum cardinality - 1 means that there must be at least one value (required), 0 means that the value is optional
  - Maximum cardinality - 1 means that there can be at most one value (single-valued), N means that there can be up to N values (N > 1, multi-valued)

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Simple number restriction examples

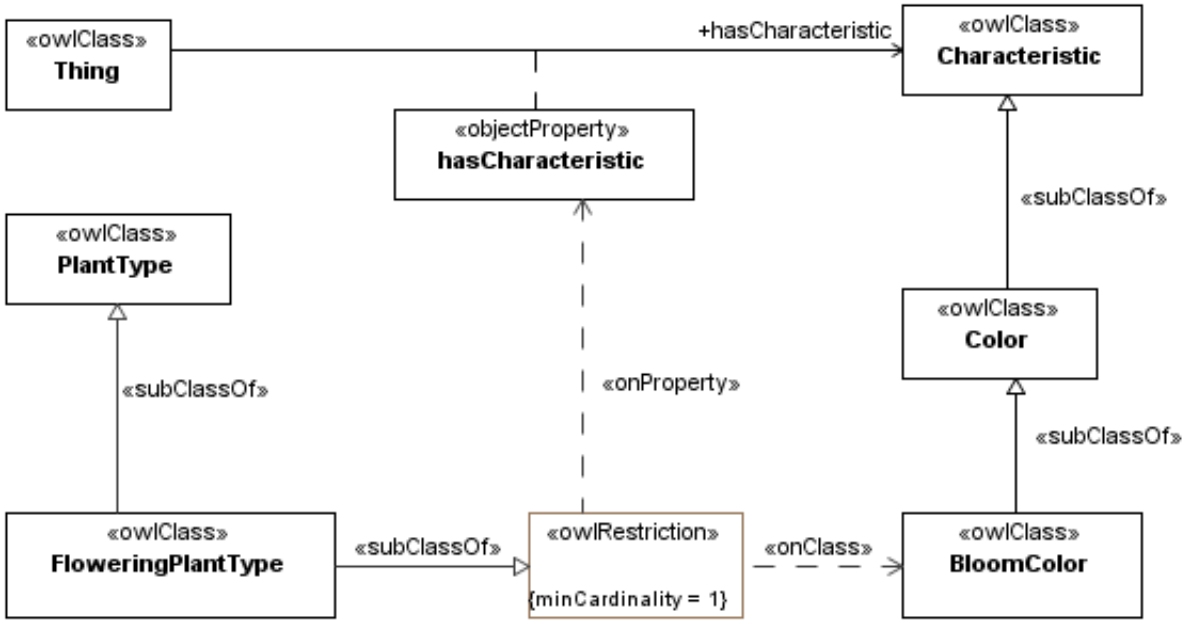



```
<owl:ObjectProperty rdf:about="&re;hasColor">
    <rdfs:range rdf:resource="&re;Color"/>
    <rdfs:domain rdf:resource="&re;ColoredThing"/>
  </owl:ObjectProperty>

  <owl:Class rdf:about="&re;Color"/>
  <owl:Class rdf:about="&re;ColoredThing"/>

  <owl:Class rdf:about="&re;MultiColoredThing">
    <owl:equivalentClass>
      <owl:Restriction>
        <owl:onProperty rdf:resource="&re;hasColor"/>
        <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">2</owl:minCardinality>
      </owl:Restriction>
    </owl:equivalentClass>
  </owl:Class>

  <owl:Class rdf:about="&re;SingleColoredThing">
    <owl:equivalentClass>
      <owl:Restriction>
        <owl:onProperty rdf:resource="&re;hasColor"/>
        <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
      </owl:Restriction>
    </owl:equivalentClass>
  </owl:Class>
```

- Creating a class of individuals that have exactly one color & a class of individuals that have more than one color
- Note that the relationship between SingleColoredThing and the restriction class is modeled as an *equivalence relationship*, meaning that membership in the restriction class is a necessary and sufficient condition for being a SingleColoredThing

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Qualified cardinality number restriction example



- OWL 2 provides the capability to further qualify cardinality by specific classes and data ranges
- Patterns that reuse a smaller number of significant properties, building up more complex class expressions that limit the set of valid values are common, useful in complex classification systems

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Qualified cardinality number restriction example


```
<owl:ObjectProperty rdf:about="&re;hasCharacteristic">
    <rdfs:range rdf:resource="&re;Characteristic"/>
  </owl:ObjectProperty>

  <owl:Class rdf:about="&re;BloomColor">
    <rdfs:subClassOf rdf:resource="&re;Characteristic"/>
    <rdfs:subClassOf rdf:resource="&re;Color"/>
  </owl:Class>

  <owl:Class rdf:about="&re;Characteristic"/>

  <owl:Class rdf:about="&re;Color"/>

  <owl:Class rdf:about="&re;FloweringPlantType">
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="&re;hasCharacteristic"/>
        <owl:onClass rdf:resource="&re;BloomColor"/>
        <owl:minQualifiedCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:minQualifiedCardinality>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
```

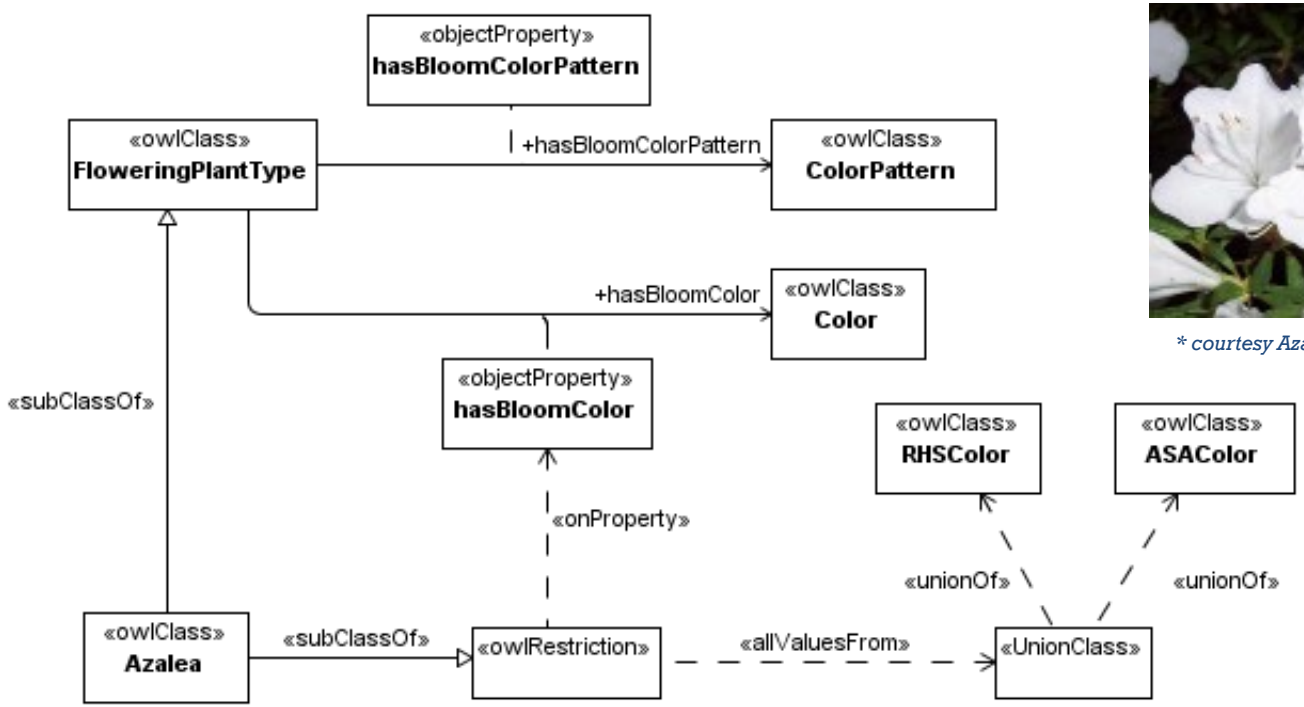
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Class expressions that restrict possible values by type

- Universal (*allValuesFrom*) and existential (*someValuesFrom*) quantification
  - ***allValuesFrom*** is the OWL equivalent for ***∀(x)***, and restricts the possible values for a property to members of a particular class or data range
  - ***someValuesFrom*** is the OWL equivalent for ***∃(x)***, and restricts the possible values for a property to at least one member of a particular class or data range
- Specific value (***hasValue***) restrictions
  - a single data value (*e.g.*, the color property for a RedWine must be filled with the value "red")
  - an individual member of a class (*e.g.*, Winery is the value restriction on the hasMaker property on the class Wine)
- Enumerations – lists of allowable individuals or data elements
- Combinations of the above that include both restrictions and boolean class expressions (union – *logical or*, intersection – *logical and*, complement – *logical not*)

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Class expression restricting *all values* for a given property



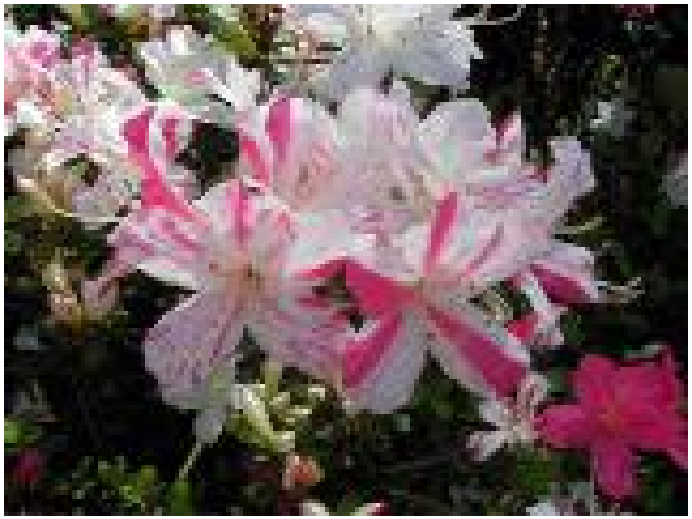
** courtesy Azalea Society of America*

- Azaleas are members of the set of flowering plants whose bloom color must be either an RHSColor (Royal Horticultural Society) or ASAColor (Azalea Society of America)

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

And the OWL for that …

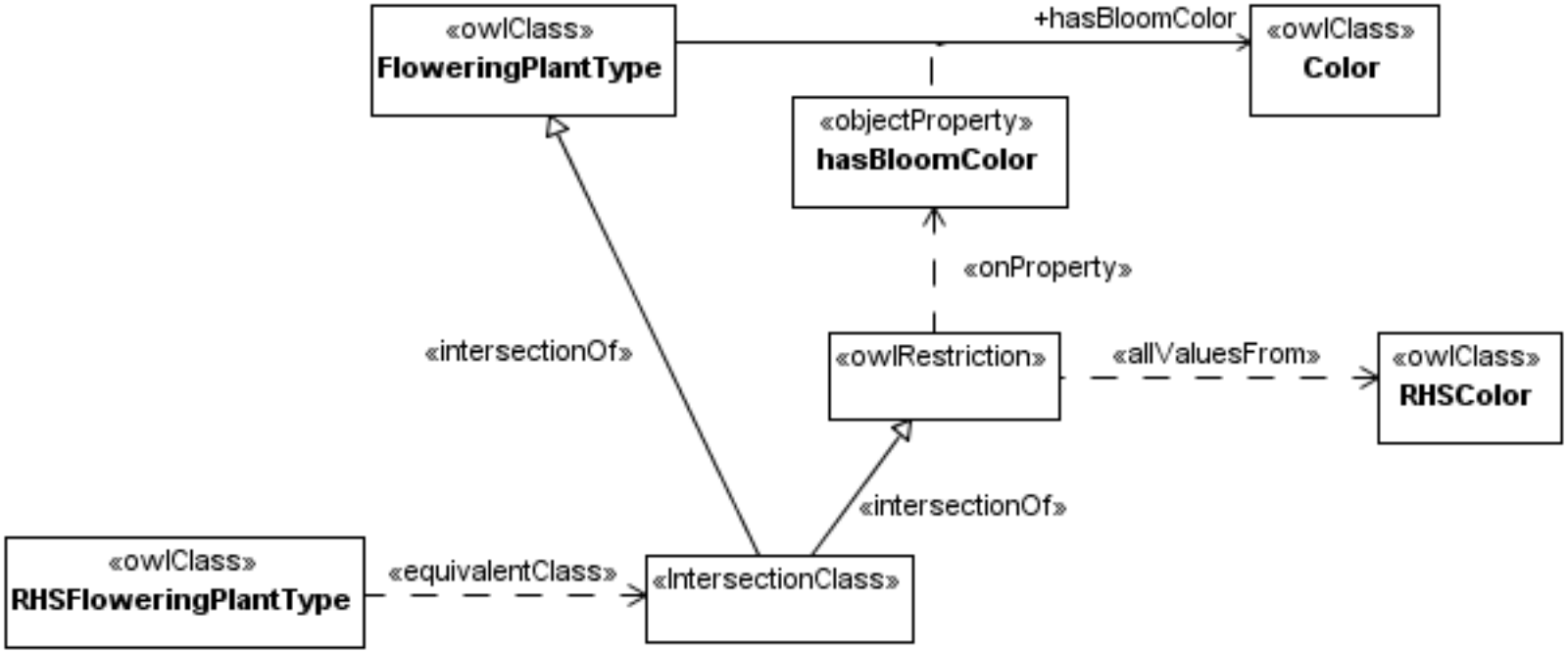

```
<owl:Class rdf:about="&re;Azalea">
  <rdfs:subClassOf rdf:resource="&re;FloweringPlantType"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="&re;hasBloomColor"/>
      <owl:allValuesFrom>
        <owl:Class>
          <owl:unionOf rdf:parseType="Collection">
            <rdf:Description rdf:about="&re;ASAColor"/>
            <rdf:Description rdf:about="&re;RHSColor"/>
          </owl:unionOf>
        </owl:Class>
      </owl:allValuesFrom>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```



** courtesy Azalea Society of America*

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Another typical pattern combining restrictions & Boolean connectives



- Use of equivalence expressions like this, describing RHSFloweringPlantType as a flowering plant and something whose bloom colors must be from the set of colors defined by the Royal Horticultural Society is a common pattern

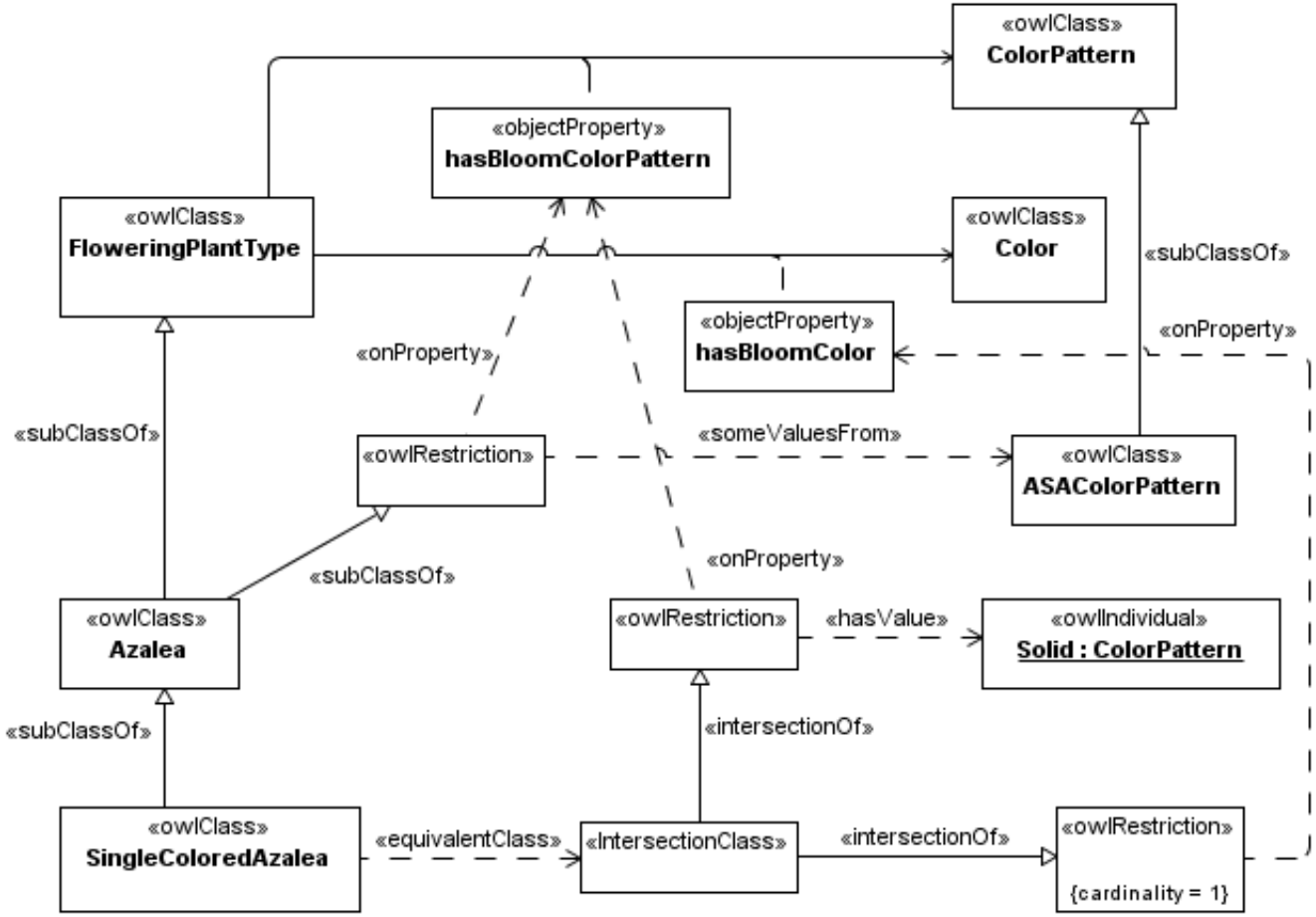
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

And the corresponding OWL for that …

```
<owl:Class rdf:about="&re;RHSFloweringPlantType">
    <owl:equivalentClass>
      <owl:Class>
        <owl:intersectionOf rdf:parseType="Collection">
          <rdf:Description
rdf:about="&re;FloweringPlantType"/>
          <owl:Restriction>
            <owl:onProperty
rdf:resource="&re;hasBloomColor"/>
            <owl:allValuesFrom
rdf:resource="&re;RHSColor"/>
          </owl:Restriction>
        </owl:intersectionOf>
      </owl:Class>
    </owl:equivalentClass>
  </owl:Class>
```

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Class expression restricting some & exact values for a given property



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

And the OWL for that …


```
<owl:Class rdf:about="&re;Azalea">
    <rdfs:subClassOf rdf:resource="&re;FloweringPlantType"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="&re;hasBloomColor"/>
        <owl:allValuesFrom>
          <owl:Class>
            <owl:unionOf rdf:parseType="Collection">
              <rdf:Description rdf:about="&re;ASAColor"/>
              <rdf:Description rdf:about="&re;RHSColor"/>
            </owl:unionOf>
          </owl:Class>
        </owl:allValuesFrom>
      </owl:Restriction>
    </rdfs:subClassOf>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="&re;hasBloomColorPattern"/>
        <owl:someValuesFrom rdf:resource="&re;ASAColorPattern"/>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
```

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

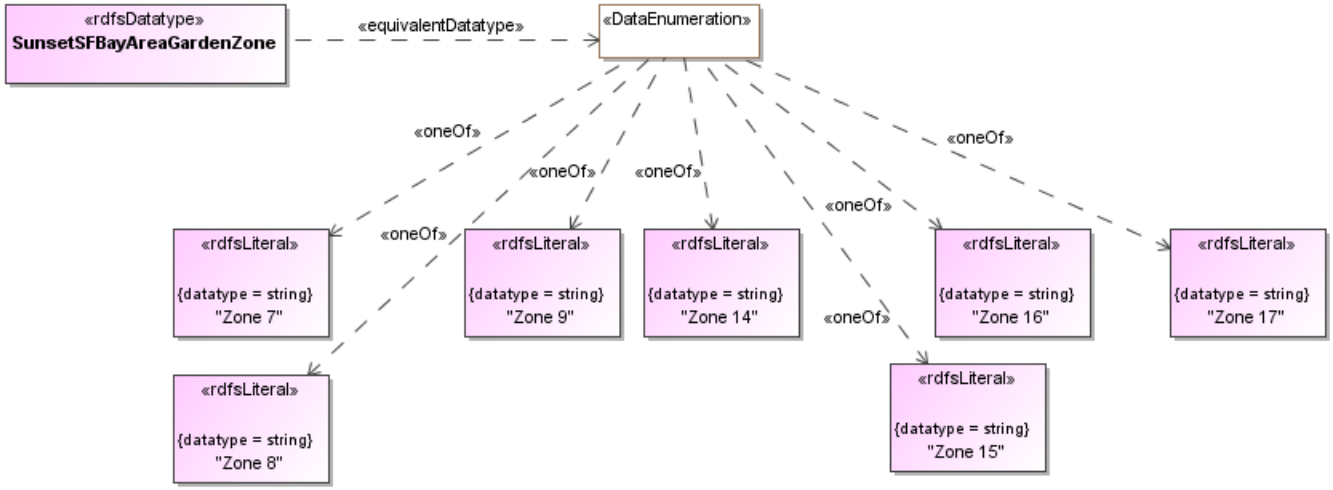
...

```
<owl:Class rdf:about="&re;SingleColoredAzalea">
    <owl:equivalentClass>
      <owl:Class>
        <owl:intersectionOf rdf:parseType="Collection">
          <owl:Restriction>
            <owl:onProperty rdf:resource="&re;hasBloomColorPattern"/>
            <owl:hasValue rdf:resource="&re;Solid"/>
          </owl:Restriction>
          <owl:Restriction>
            <owl:onProperty rdf:resource="&re;hasBloomColor"/>
            <owl:cardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
          </owl:Restriction>
        </owl:intersectionOf>
      </owl:Class>
    </owl:equivalentClass>
    <rdfs:subClassOf rdf:resource="&re;Azalea"/>
  </owl:Class>

  <owl:NamedIndividual rdf:about="&re;Solid">
    <rdf:type rdf:resource="&re;ColorPattern"/>
  </owl:NamedIndividual>
```

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Individuals and data ranges

- An individual (instance, object in other paradigms)
  - Any class that an individual *is a member of*, or *is an individual of*, is a type of the individual
  - Any superclass of a class *is an ancestor of* (*or type of*) the individual
- Specify property values for the individual
  - Property values should conform to the constraints such as range, value type, cardinality restrictions, etc.
- Enumerated classes & data ranges are commonly used to specify the complete set of valid values for a particular property



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Example enumerated class of individuals

RestrictionExamples (http://www.omg.org/spec/ODM/20120424/examples/RestrictionExamples#RestrictionExamples) - [C:\Users\ekendall\VOM\export\http\www.omg.org\spec\ODM\20120424\examples\Restriction...

File Edit View Reasoner Tools Refactor Window Help

RestrictionExamples (http://www.omg.org/spec/ODM/20120424/examples/RestrictionExamples#RestrictionExamples)

Active Ontology | Entities | Classes | Object Properties | Data Properties | Individuals | OWLViz | DL Query | OntoGraf

Class hierarchy (inferred) | Class hierarchy

Class hierarchy (inferred): Colo

- Thing
  - BrightColor
  - BrightColoredThing
  - Characteristic
  - Color
  - ColorPattern
  - ColoredThing
  - FlowerForm
  - FloweringPlantType
  - FloweringPlantType'
  - PetalForm

Members list (inferred) | Members list

Members list: Solid

- Blotched
- Center
- Dash
- Edge
- Line
- Marble
- ShadeTo
- Solid
- Speckle
- Spot
- Streak
- Stripe
- Suffused
- Tinge
- Variegated
- Veined

Annotations | Usage

Usage: Solid

Show: this different

Found 4 uses of Solid

- MultiColoredAzalea
  - MultiColoredAzalea SubClassOf Azalea and MultiColoredThing and (not (hasBloomColorPattern value Solid))
- SingleColoredAzalea
  - SingleColoredAzalea EquivalentTo (hasBloomColorPattern value Solid) and (hasBloomColor exactly 1 Thing)
- Solid
  - Solid Type ColorPattern
  - Individual: Solid

Description: Solid

Types

ColorPattern

Same individuals

Different individuals

Property assertions: Solid

Object property assertions

Data property assertions

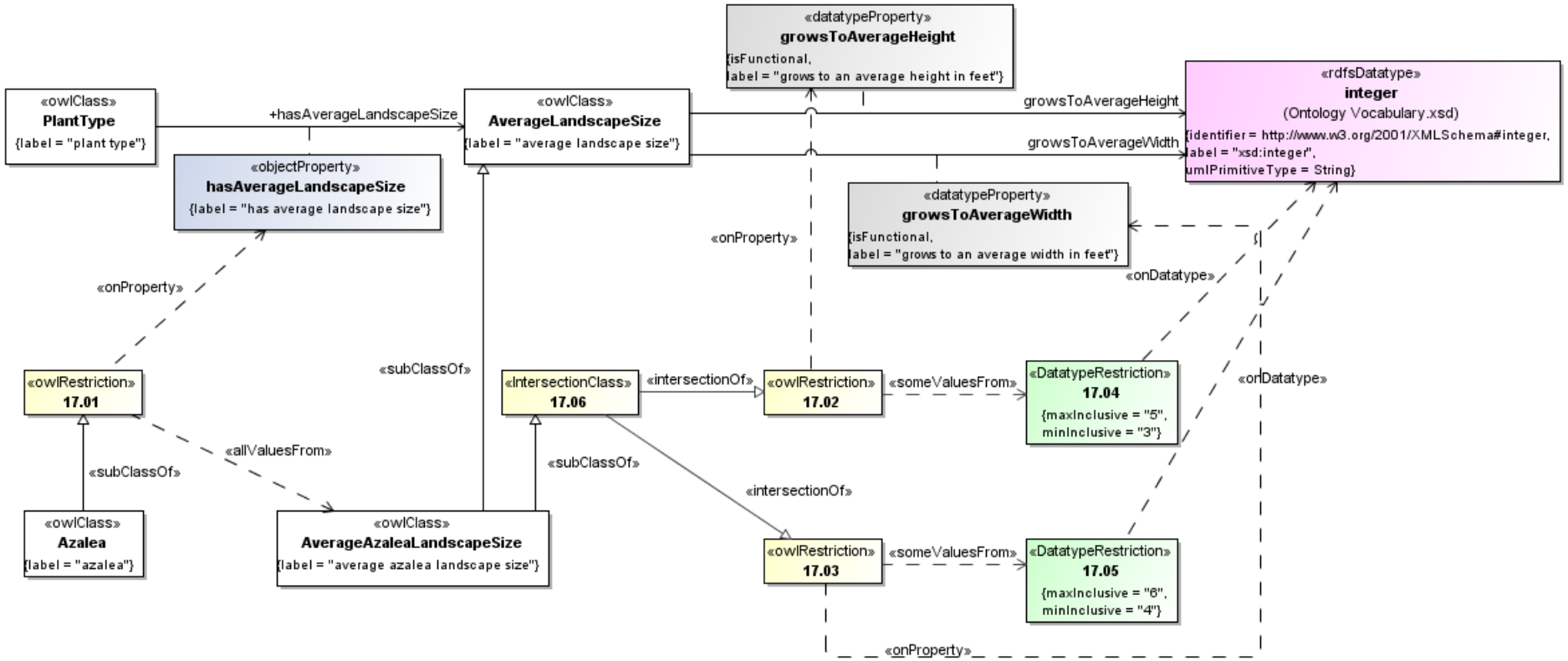
Negative object property assertions

Negative data property assertions

Reasoner active | Show Inferences

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Complex data ranges & restrictions



Aggregating datatype restrictions via an intersection to specify the average size of an Azalea in the landscape.

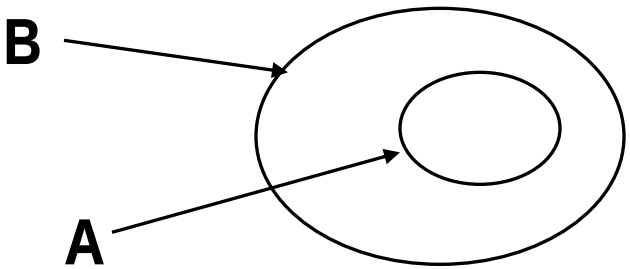
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Class axioms

- Subsumption (necessary)
  - $A \subseteq B$ where B

  is a class description

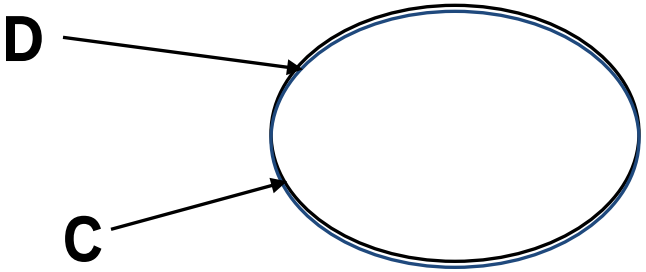
  partial or *primitive class*



- Definition (necessary and sufficient)
  - $C \equiv D$ where D

  is a class description

  complete or *defined class*

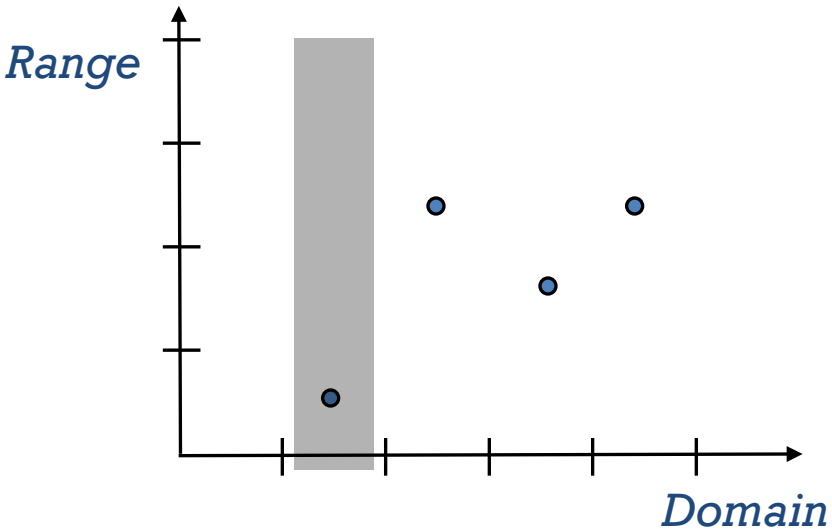


Courtesy Evan Wallace, NIST

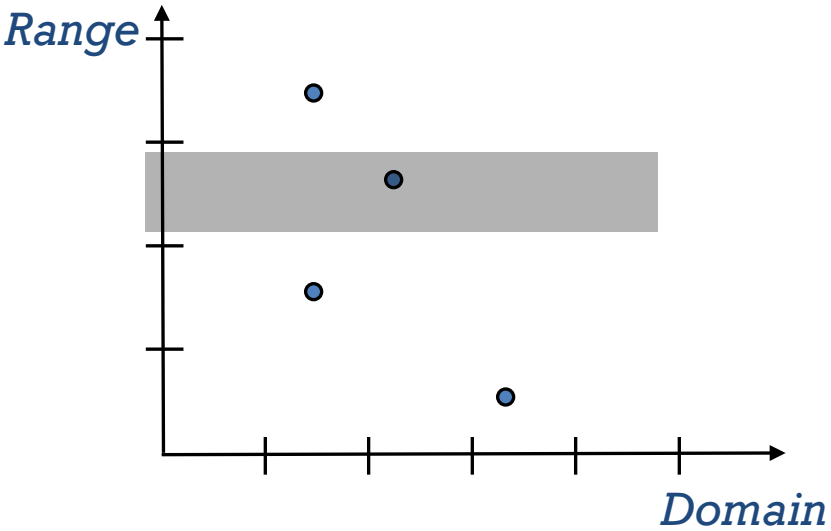
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Meta-properties – global cardinality restrictions

- Functional



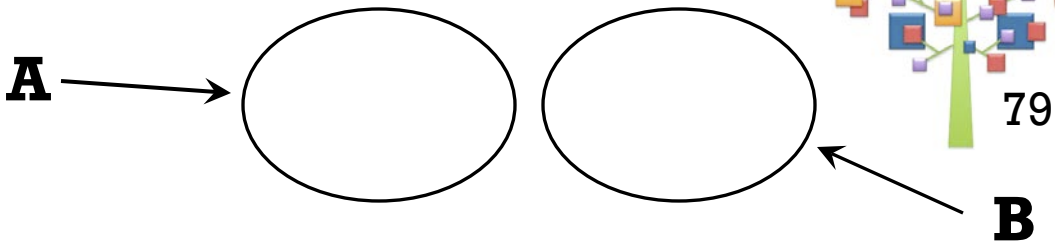
- Inverse Functional



Courtesy Evan Wallace, NIST

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Disjoint classes



- Classes are disjoint if they cannot have common instances
- Disjoint classes cannot have any common subclasses
- If winery and wine are disjoint, then there is no instance that is both a winery and a wine; there is no class that is both a subclass of winery and a subclass of wine
- Disjointness is often used to aid consistency checking
- Disjointness is also helpful in teasing out subtle distinctions among classes across multiple ontologies
- Equivalence is also often used to identify the same concepts across ontologies that may be named differently, or to name classes defined through class axioms

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Advanced OWL 2 language features

- Property constructs
  - Reflexive, irreflexive, & asymmetric properties
  - Property chains – useful for building up complex properties for
    - ownership and control in business entity relationships for finance
    - complex family relationships (*e.g.*, uncleOf)
  - Disjoint properties
  - Keys – for transformations to logical data models
  - Self-reflexive properties – such as narcissism
  - Negative property assertions

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

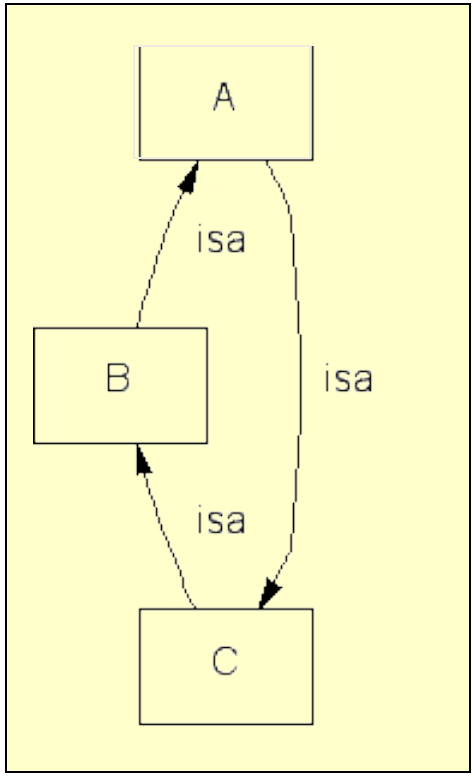
More advanced language features

- Extended datatype handling
  - Better support for XML Schema Datatypes
  - Datatypes for specifically for OWL (owl:real, owl:rational, rdf:PlainLiteral)
  - Custom datatype definition & use of xsd facets / datatype restrictions
- Data range restrictions
  - Intersections, unions, complements
- Support for punning
  - Limited support for a particular element to be defined as both a class & individual, for example
- Improved support for annotations (*e.g.*, metadata about ontologies, provenance, transformation detail, etc.)
- Related ontologies available from the W3C
  - Prov-O Provenance Ontology -- http://www.w3.org/TR/prov-overview/ and http://www.w3.org/TR/2013/REC-prov-o-20130430/
  - W3C Organization Ontology -- http://www.w3.org/TR/vocab-org/
- See http://www.w3.org/TR/owl2-quick-reference/ for OWL language details

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

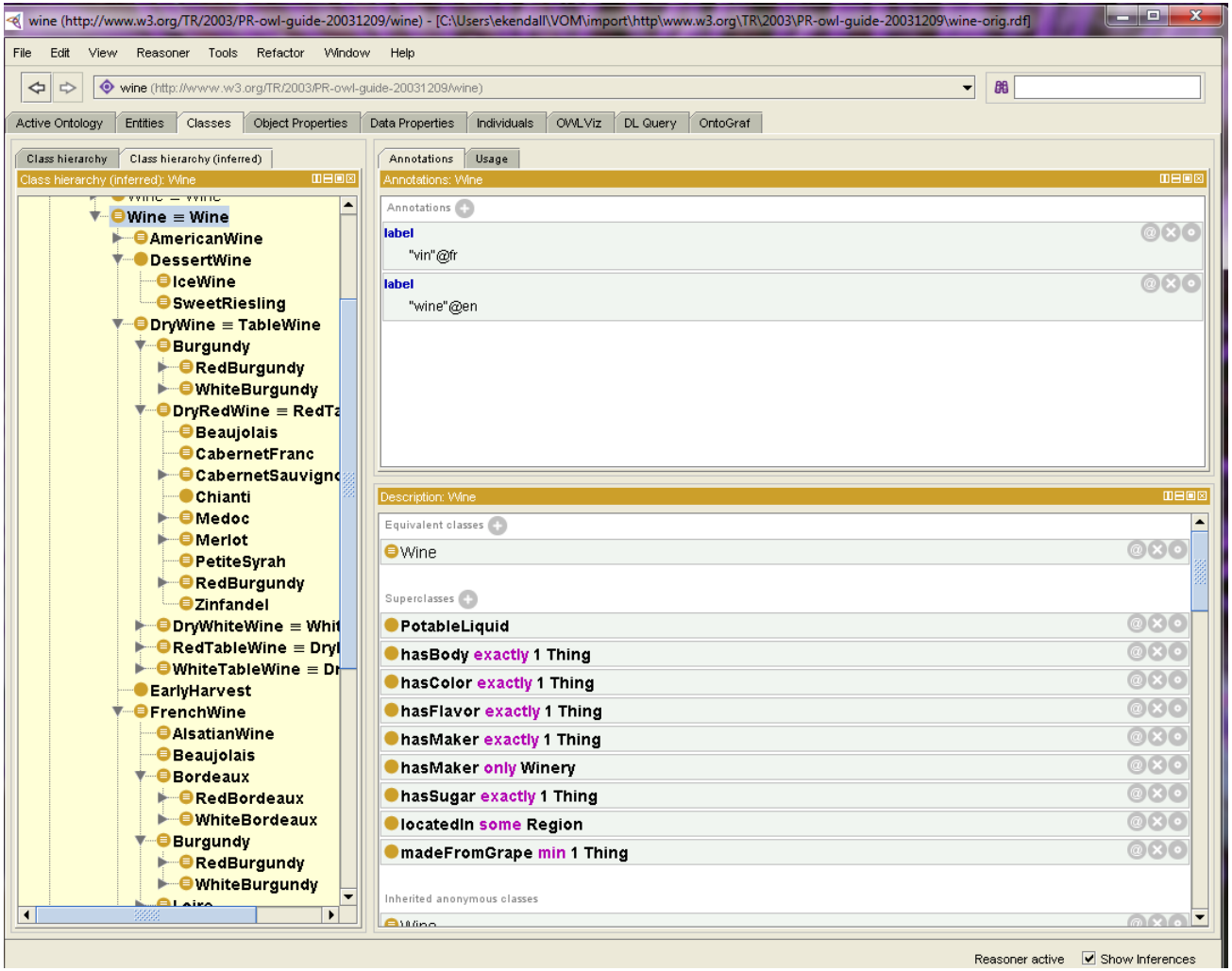
Rules of Thumb

- Cycles are common in many KR systems, though rarely "a good thing"
- Cycles are disallowed by some tools because they prohibit "code generation", export -- including RDF/OWL
- Classes A, B, and C have equivalent sets of instances
  - By many definitions, A, B, and C are equivalent
  - Use owl:equivalentClass instead of creating cycles



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Class hierarchy analysis



- Siblings should be specified at roughly the same level of generality -- compare to section and subsections in a book

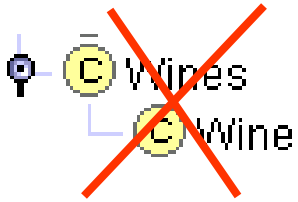
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

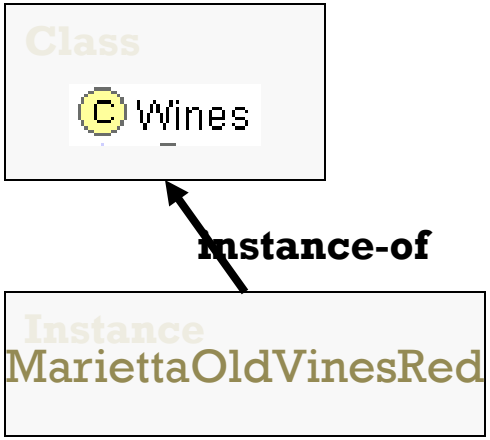
More on class definition analysis

- Classes that have single subclasses may be incompletely defined, unless additional children are specified in subordinate modules
- Subclasses of a class typically have more detailed definitions
  - Additional properties
  - Constraints/restrictions on inherited properties
  - Participate in further qualified relations
  - Compare to bullets in a bulleted list
- If a class has a large number of subclasses, it may be useful to define intermediate levels
  - For wine, for example there are natural groupings around wine color
  - If no natural classification exists, the long list may be appropriate

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Inheritance, naming, synonyms





- A "wine" is not a subclass of "wines"
- A particular vintage should be classified as an instance of the class Wines
- Class names should be either all singular or all plural
- Synonymous names for the same concept should be modeled as alternate designations, not as unique classes (in the same ontology)
- Many systems & standards facilitate synonymous term representation (*e.g.*, ISO 1087)
- OWL allows defining necessary and sufficiency condition definitions thereby allowing synonym definitions to be "first class" terms (as appropriate, through equivalence / same as relations)

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Class vs. property value

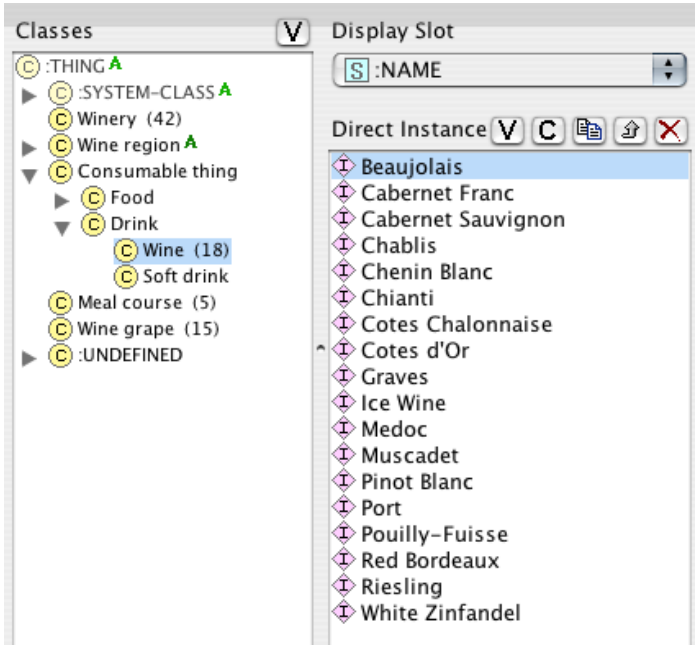
OR

- Are concepts with distinct property values used in restrictions on other properties?
- How important is the distinction for the domain?
- Class definitions for most domains should be fairly stable – *i.e.*, they should not change frequently once the definitions are established and individuals created

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Class vs. individual

OR



- Individuals are the most specific entities in an ontology
- If concepts form a natural hierarchy, represent them as classes
- If they might have individuals, represent them as classes

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Limiting scope

- An ontology should not contain all the possible information about the domain
  - No need to specialize or generalize more than the application requires
  - No need to include all possible properties of a class
    - Only the most salient properties
    - Only the properties that the application requires
- Ontologies of wine, food, and their pairings probably will not include details such as:
  - Bottle size (half bottle, full bottle, magnum, …)
  - Label color
  - Wine bottle color (green, amber, …)
  - Individual plants and their location in a historic garden
  - Azalea indica 'Alaska' (Rutherfordiana hybrid) might be a class or might be an individual, depending on the use case and application requirements

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Part 3: Putting It All Together

- Ontology-powered applications (both light weight and richer applications)
  - Semantic Advising (Wine Agent)
  - Semantically-enabled Environmental Monitoring – SemantEco and Jefferson Project
  - Population Science Health Data Exploration
  - Scientific Observations and Virtual Observatories
- Technical Underpinnings
  - Understanding, Trust, & Proof: InferenceWeb
  - Semantic Methodology
  - Linked Data
  - Towards future semantic cognitive assistants
- Questions

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

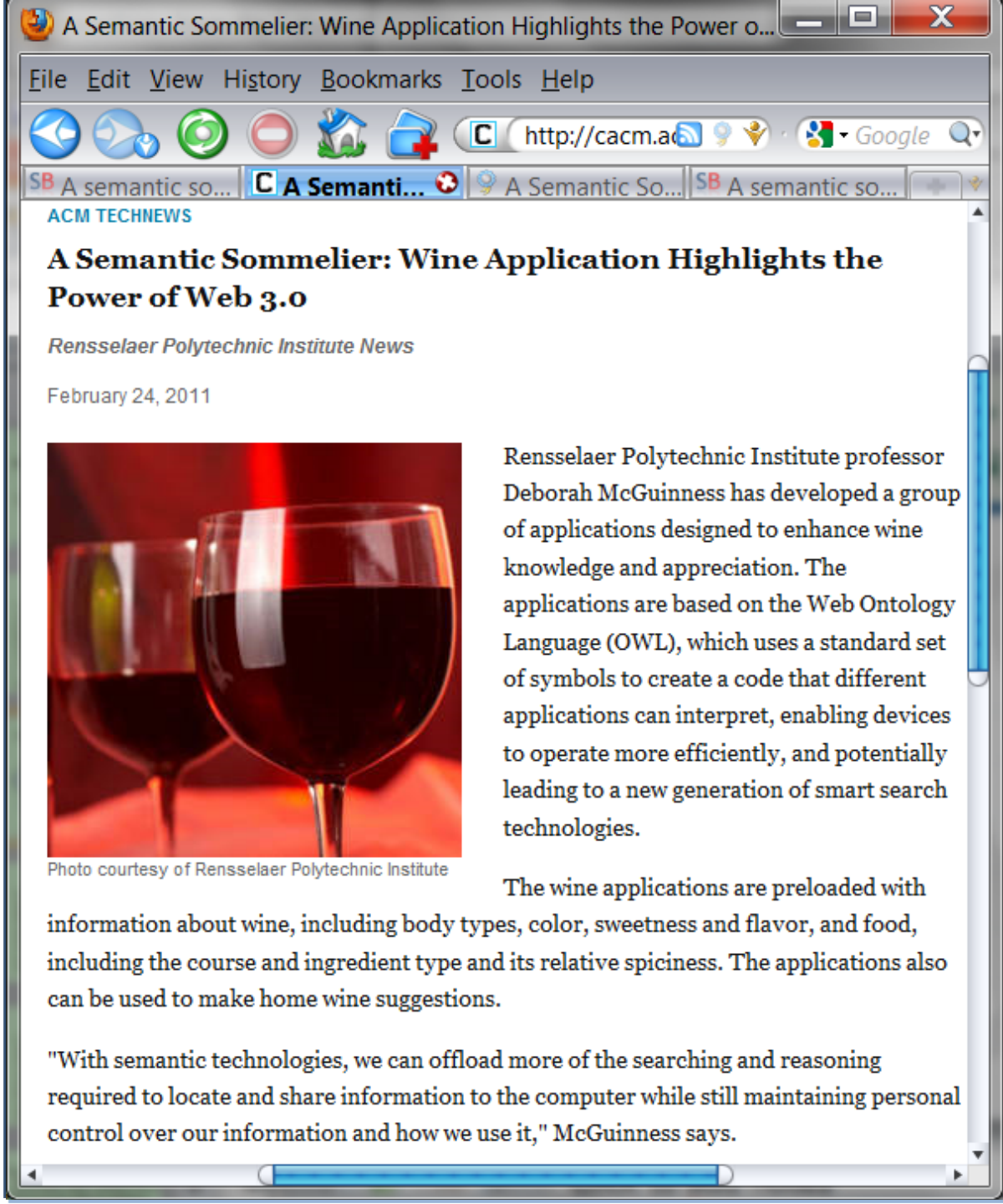
Semantic Sommelier… ontology-driven, context-aware, mobile app

Semantically-enabled advisors utilize:

- Ontologies
- Reasoning
- Social
- Mobile
- Provenance
- Context

Patton & McGuinness.et. al

tw.rpi.edu/web/project/Wineagent



A Semantic Sommelier: Wine Application Highlights the Power o...

File Edit View History Bookmarks Tools Help

http://cacm.ac

A semantic so... | A Semanti... | A Semantic So... | A semantic so...

ACM TECHNEWS

A Semantic Sommelier: Wine Application Highlights the Power of Web 3.0

Rensselaer Polytechnic Institute News

February 24, 2011

Photo courtesy of Rensselaer Polytechnic Institute

Rensselaer Polytechnic Institute professor Deborah McGuinness has developed a group of applications designed to enhance wine knowledge and appreciation. The applications are based on the Web Ontology Language (OWL), which uses a standard set of symbols to create a code that different applications can interpret, enabling devices to operate more efficiently, and potentially leading to a new generation of smart search technologies.

The wine applications are preloaded with information about wine, including body types, color, sweetness and flavor, and food, including the course and ingredient type and its relative spiciness. The applications also can be used to make home wine suggestions.

"With semantic technologies, we can offload more of the searching and reasoning required to locate and share information to the computer while still maintaining personal control over our information and how we use it," McGuinness says.

TetherlessWorld

TW Wine Agent – semantic interoperability

Wine Agent receives a meal description and retrieves a selection of matching wines available on the Web, using an ensemble of standards and tools

- RDF & OWL – for encoding wine/food listings and pairing recommendations
- Collaborative environment input (initially Semantic MediaWiki) for publishing user-contributed recommendations*
- Social Media (Twitter and Facebook) – for posting recommendations
- Pellet – for deriving knowledge from wine ontology and recommendations
- SPARQL – for querying wine/food listings with recommendations
- InferenceWeb – for explaining Wine Agent's recommendations

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Semantic Sommelier

- Uses ontologies to infer descriptions of wines for meals and query for wines
- Uses context: GPS location, local restaurants and wine lists, user preferences
- Uses Social input: Twitter, Facebook, Wiki, mobile, …
- Challenges include source variability in quality, contradictions exist, maintenance
- Customization & Extensions from restaurant, wine seller, user easy
- Extensions to reasoning like group recommendations are flexible

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Reflections from the Semantic Advisors

- Background knowledge can be reasonably simple and built in OWL. For the wine agent; includes foods and wine and pairing information similar to the original Living with CLASSIC paper, CLASSIC tutorial, OWL Guide, Ontology Engineering 101, …) and has been extended easily and regularly
- Background knowledge can be used for simple query expansion. In the wine agent, e.g., over wine sources to retrieve documents about red wines (including zinfandel, syrah, …)
- Background knowledge used to interact with structured queries such as those possible on wine sites like wine.com or apps like delectable
- Constraints allows a reasoner like Pellet to infer consequences of the premises and query.
- Explanation system (e.g., Inference Web) can provide provenance information such as information on the knowledge source (McGuinness' wine ontology) and data sources (such as wine reviews or particular restaurants or wine stores)
- Services work could allow automatic "matchmaking" instead of hand coded linkages with web resources

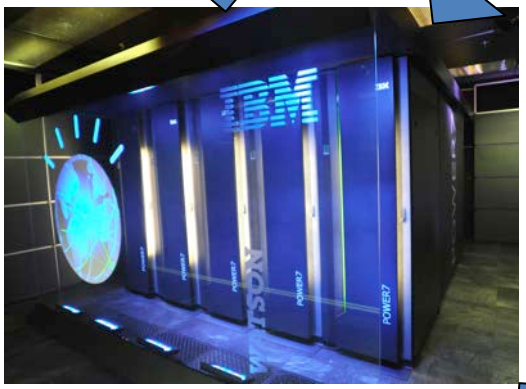
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Mobile Health Advisor

User Device

PHR/EHR

Physician



CDC Centers for Disease Control and Prevention

Icahn School of Medicine at Mount Sinai

Johns Hopkins Medicine

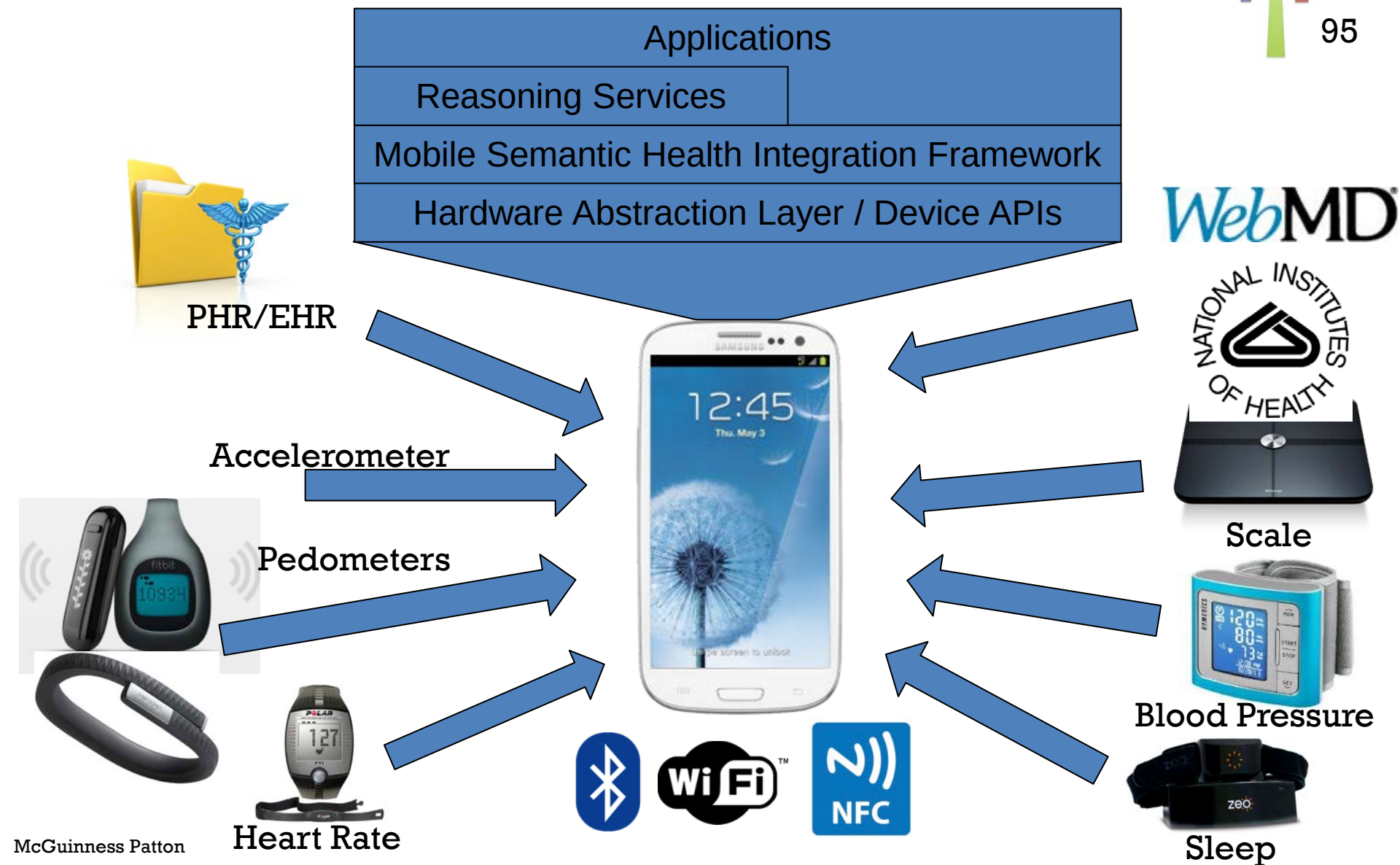
Mayo Clinic

WebMD

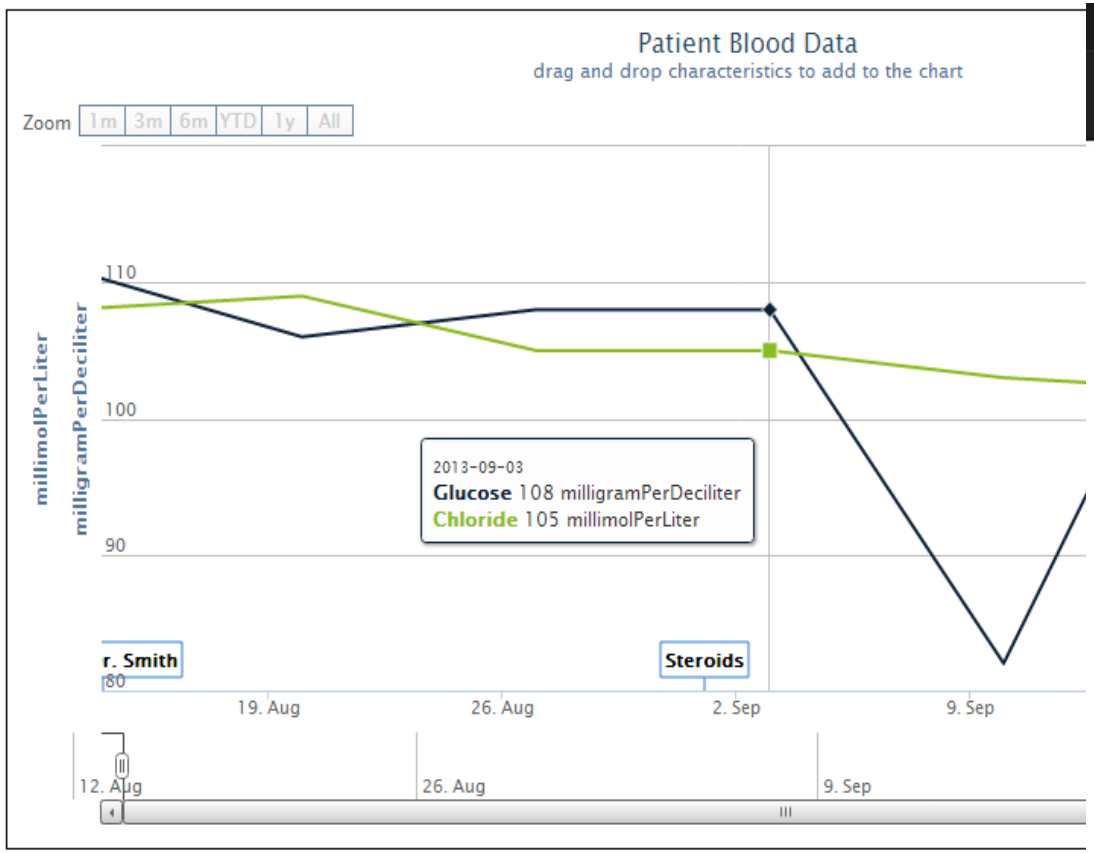
National Institutes of Health

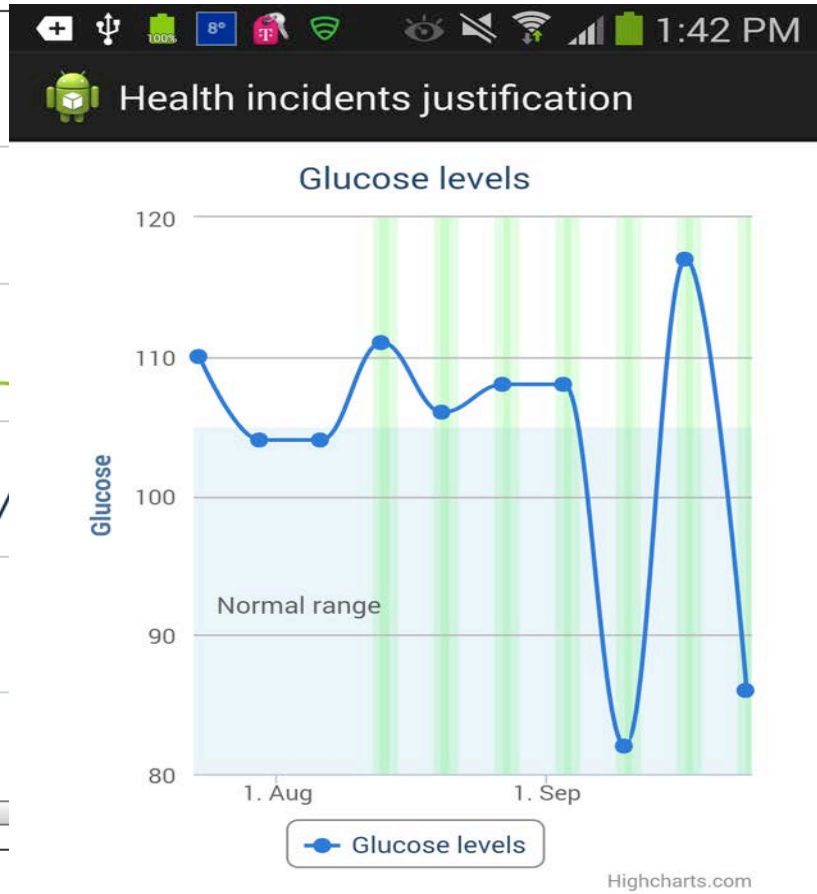
Joint work: Patton, Chastain, Makni, Ji, McGuinness

Mobile Health Monitoring



Mobile Health Monitoring





Copyright © 2015 Thematix Partners LLC & McGuinness Associates

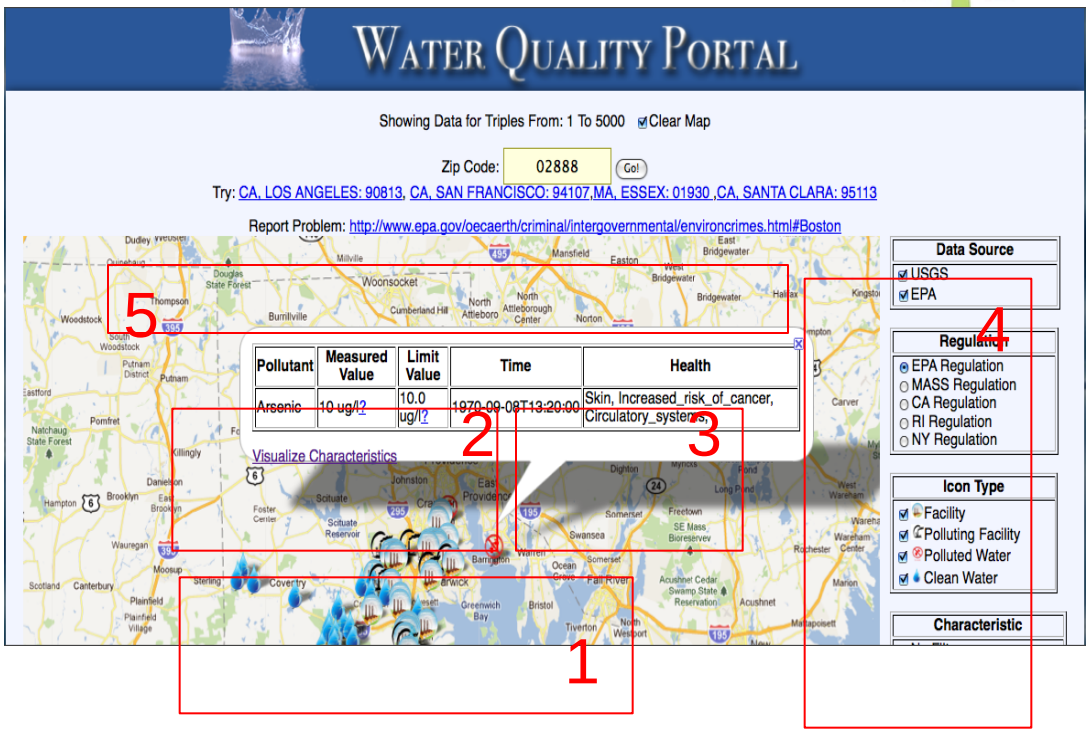
Reflections on Semantic Health Advisors

- Relatively simple background knowledge (e.g., thresholds for lab values, SOME indicators or justifications for out of range values) can be leveraged with straight forward reasoning
- Highly available resources (e.g., Drugbank, UMLS, SIDER) can be used with simple query expansion or simple subsumption.
- Linked data alone can provide significant benefits
- Provenance / explanation can be critical in some applications, particularly health.
- These demos are just scratching the surface, but the leveraging of linked data has potential to significantly improve navigation, and understanding.
- There are many more examples – in health including neural stem cell data * come to the talk this aft on Semantics meets Numerics*, ICU bed admission, child health data, etc.

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

SemantEco/SemantAqua

- Enable/Empower citizens & scientists to explore pollution sites, facilities, regulations, and health impacts along with provenance.
- Demonstrates semantic monitoring possibilities.
- Map presentation of analysis
- Explanations and Provenance available



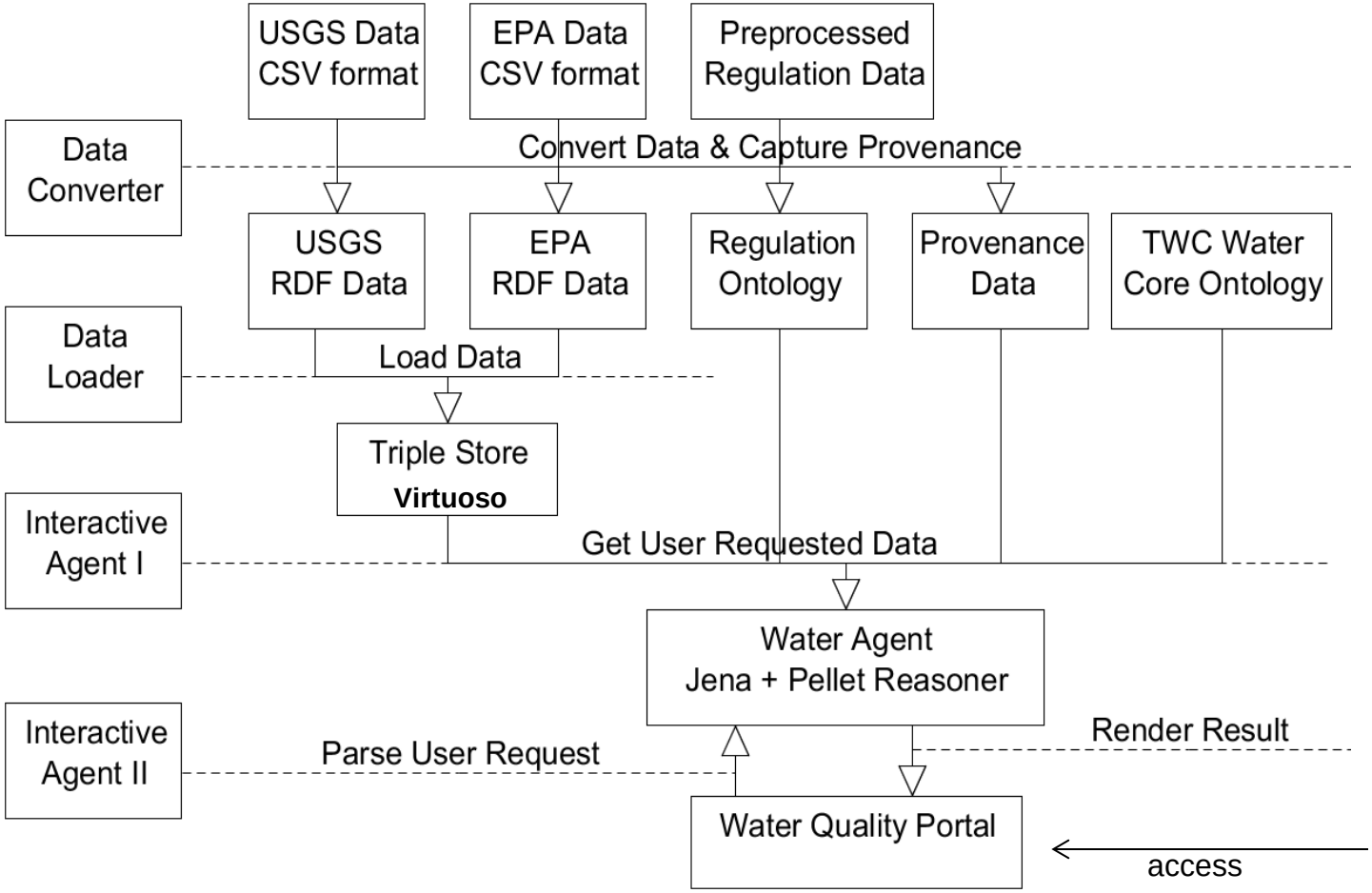
http://was.tw.rpi.edu/swqp/map.html and http://aquarius.tw.rpi.edu/projects/semantaqua

1. Map view of analyzed results
2. Explanation of pollution
3. Possible health effect of contaminant (from EPA)
4. Filtering by facet to select type of data
5. Link for reporting problems

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

System Architecture

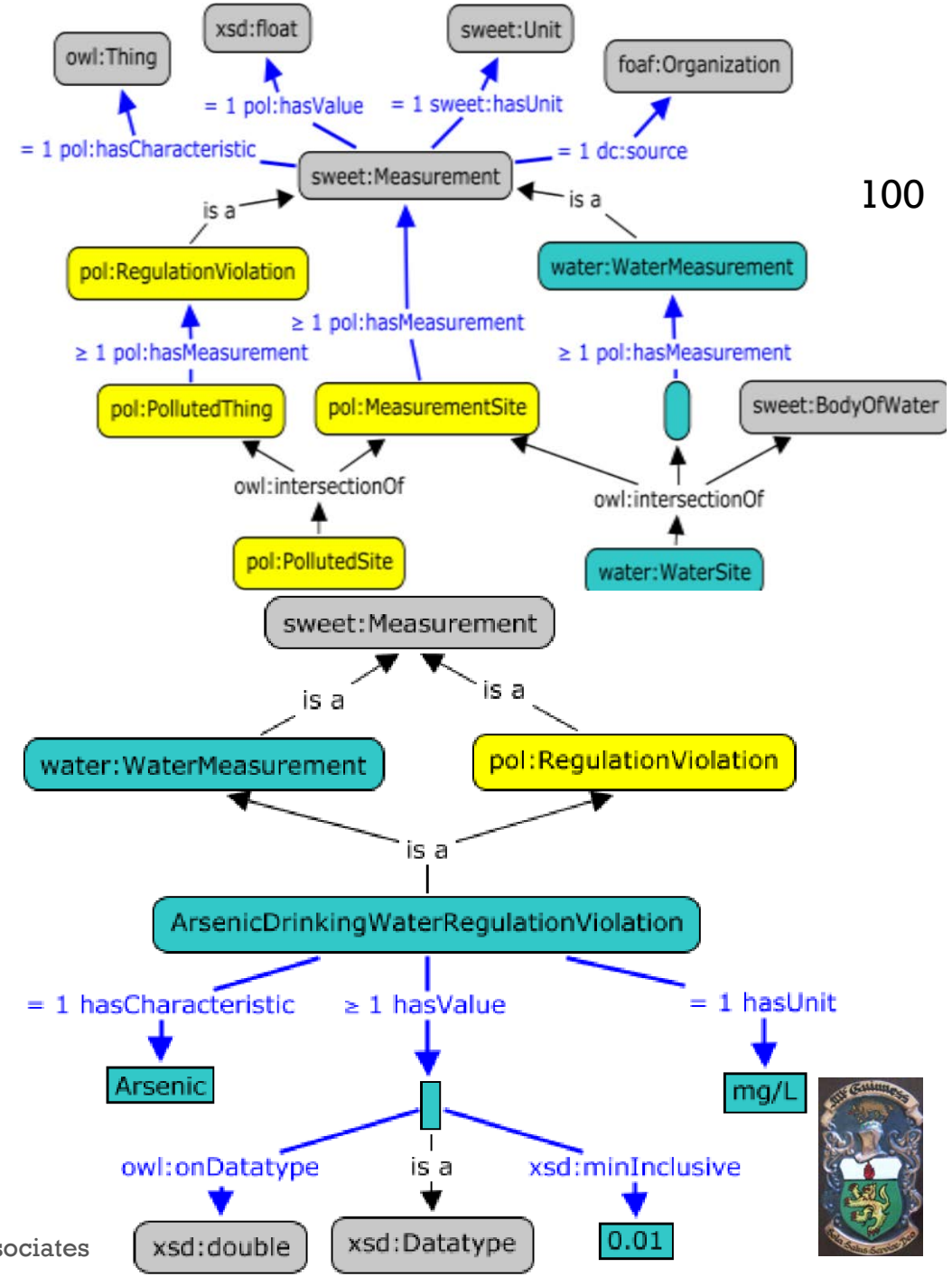
99



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Ontology

- Extends existing best practice ontologies, e.g. SWEET, OWL-Time. Includes terms for relevant pollution concepts
- Can conclude: "any water source that has a measurement outside of its allowable range" is a polluted water source.
- Regulation Ontology models federal and state water qualit regulations for drinking wate sources
- Can recognize pollution, e.g. "any water source that conta 0.01 mg/L of Arsenic or more a polluted water source."



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Querying

- With OWL reasoning during query:

```
SELECT DISTINCT ?site ?lat ?lng
WHERE { ?site a pol:PollutedSite ;
        geo:lat ?lat ; geo:long ?lng . }
```

- OWL reasoning hides the complexity of the relationships allowing the developer (or user) to ask simple questions without requiring deep knowledge of environmental regulations

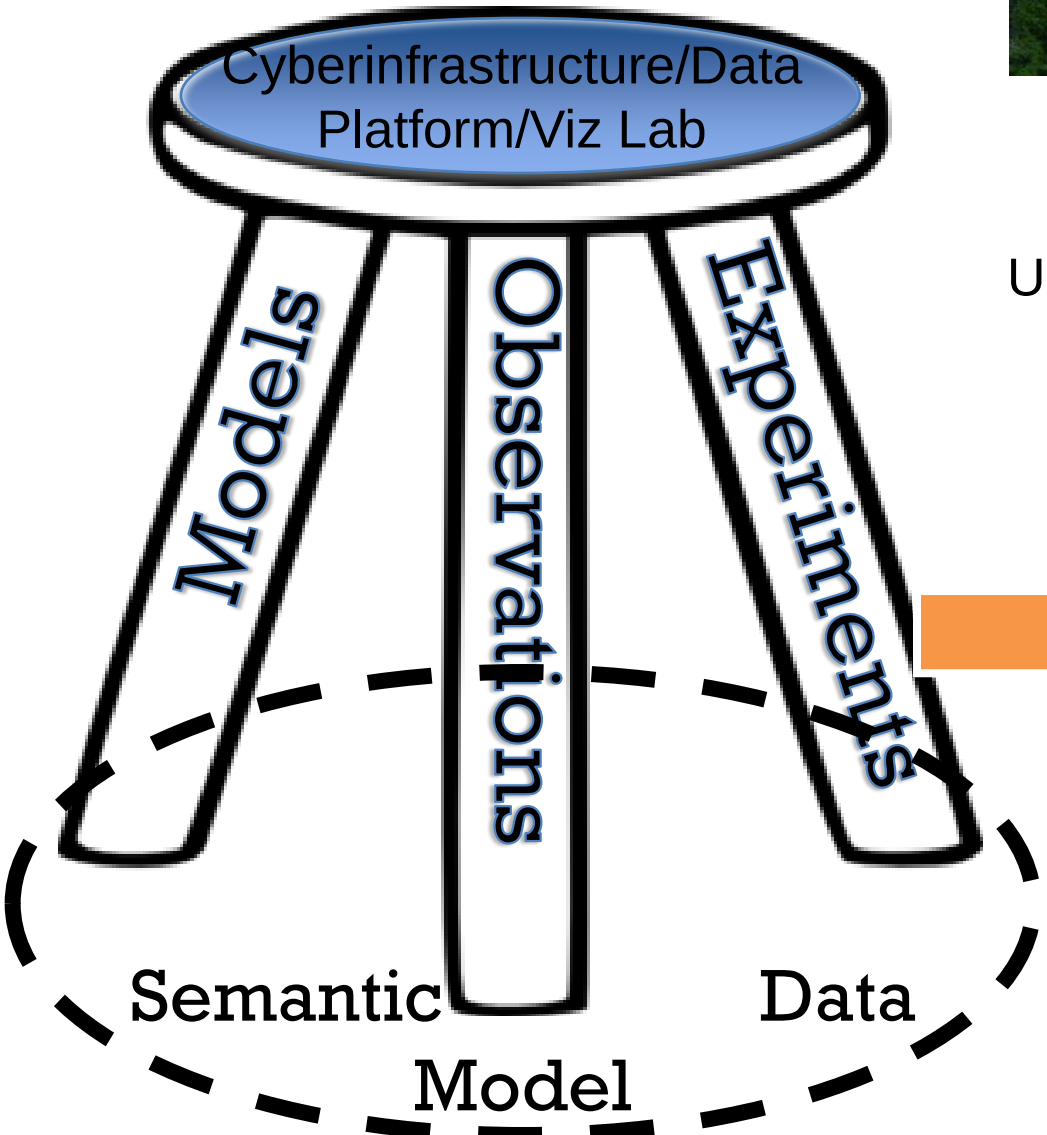
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Reflections

- Uses semantic technologies along with simple ontologies and provenance-encoding tools
- Was generated by students in a regular term class, then extended to be the basis of a masters degree (Wang), and the foundation of 3 other class projects
- Small background extensible ontology
- Has been reused in other environmental monitoring projects including content related to wildlife, migration, health effects, etc.
- Now connected with OBOE, PROV and uses HasNETO
- Forming the foundation for the Jefferson Project

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

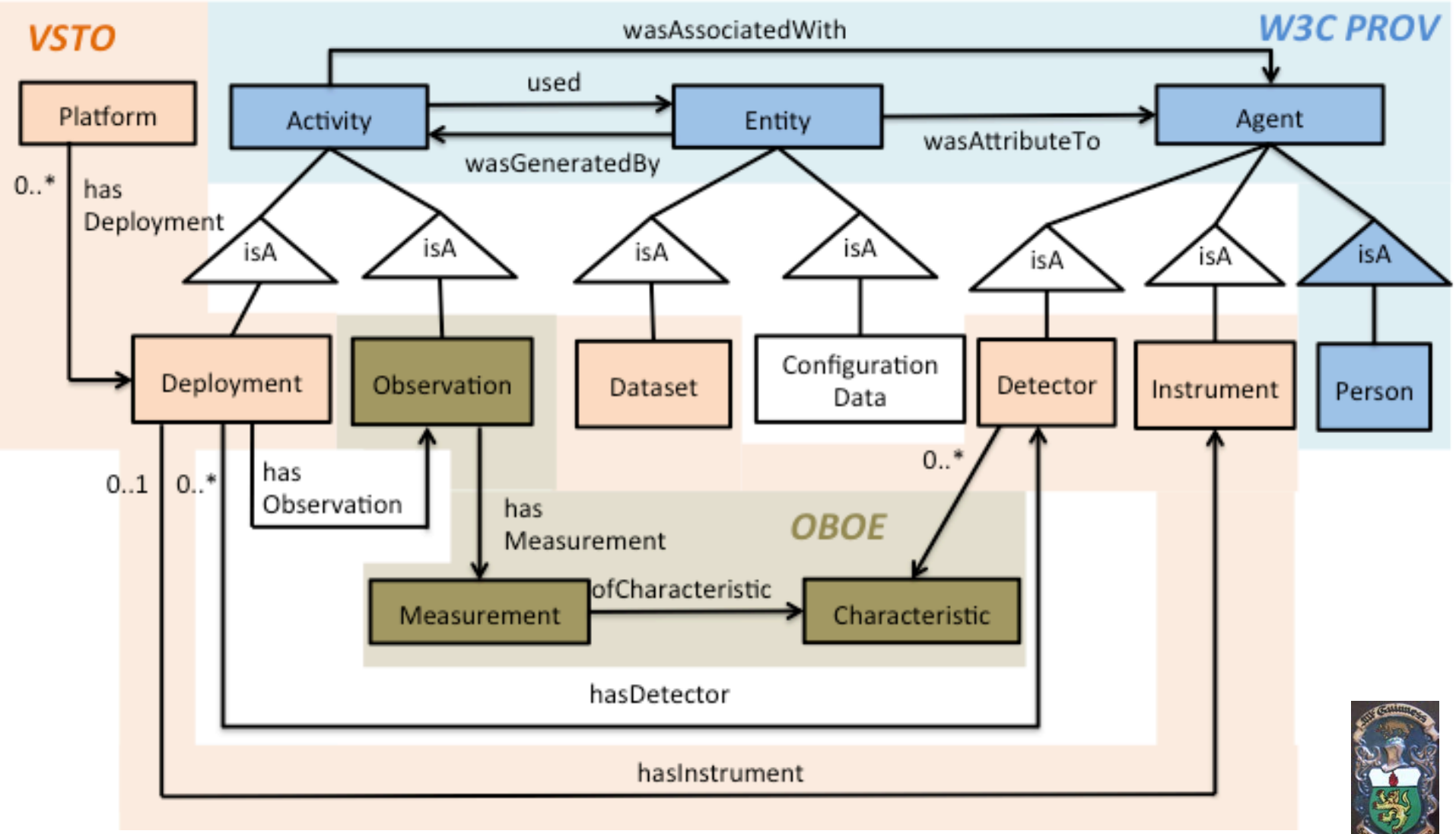
The Jefferson Project at Lake George: Science to Inform Solutions



Smart Lake:
Integrative Approach to Understanding Lake Stressors and Predicting Future Outcomes

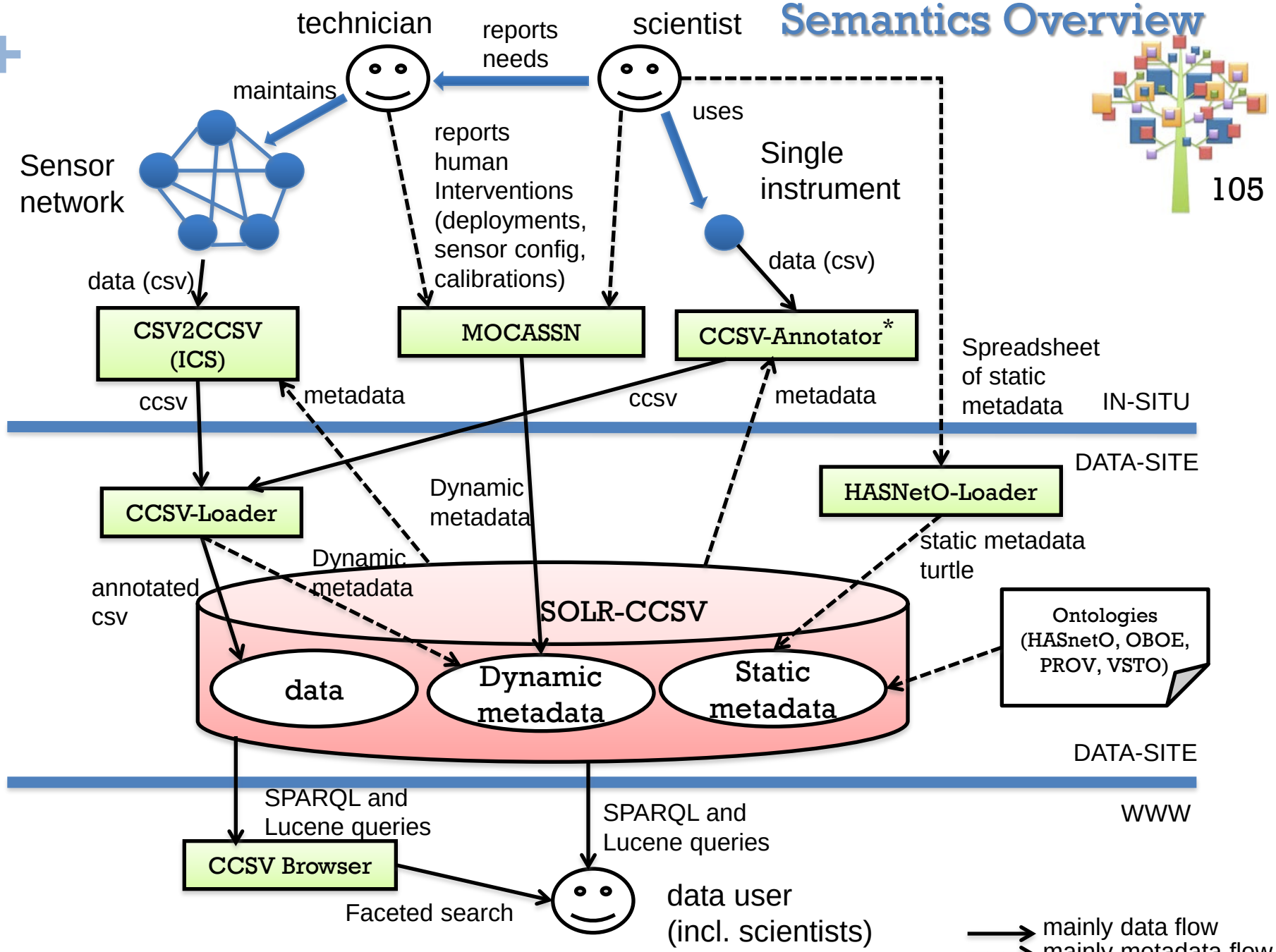
Science-based Solutions:
Leveraging deep understanding for solutions with staying power for a healthy Lake George for future generations

The Human-Aware Sensor Network Ontology



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

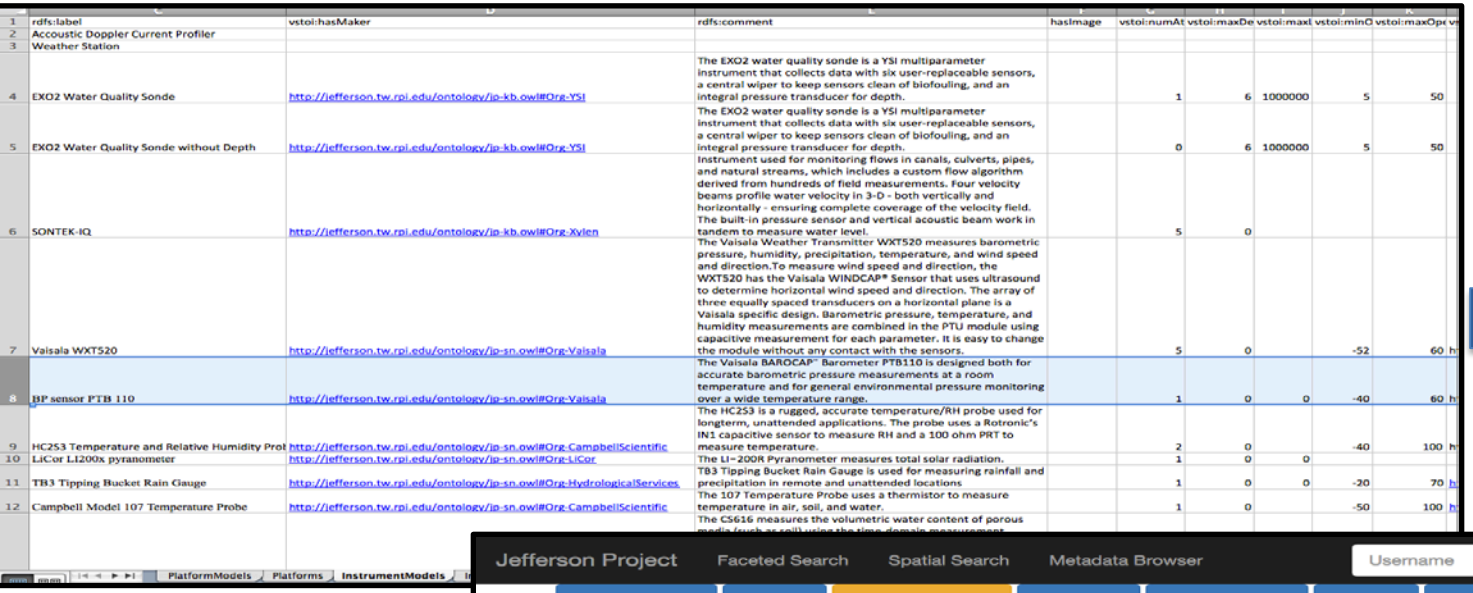
Semantics Overview



McGuinness, Pinheiro, Santos, Chastain, Kinkead, Klawonn

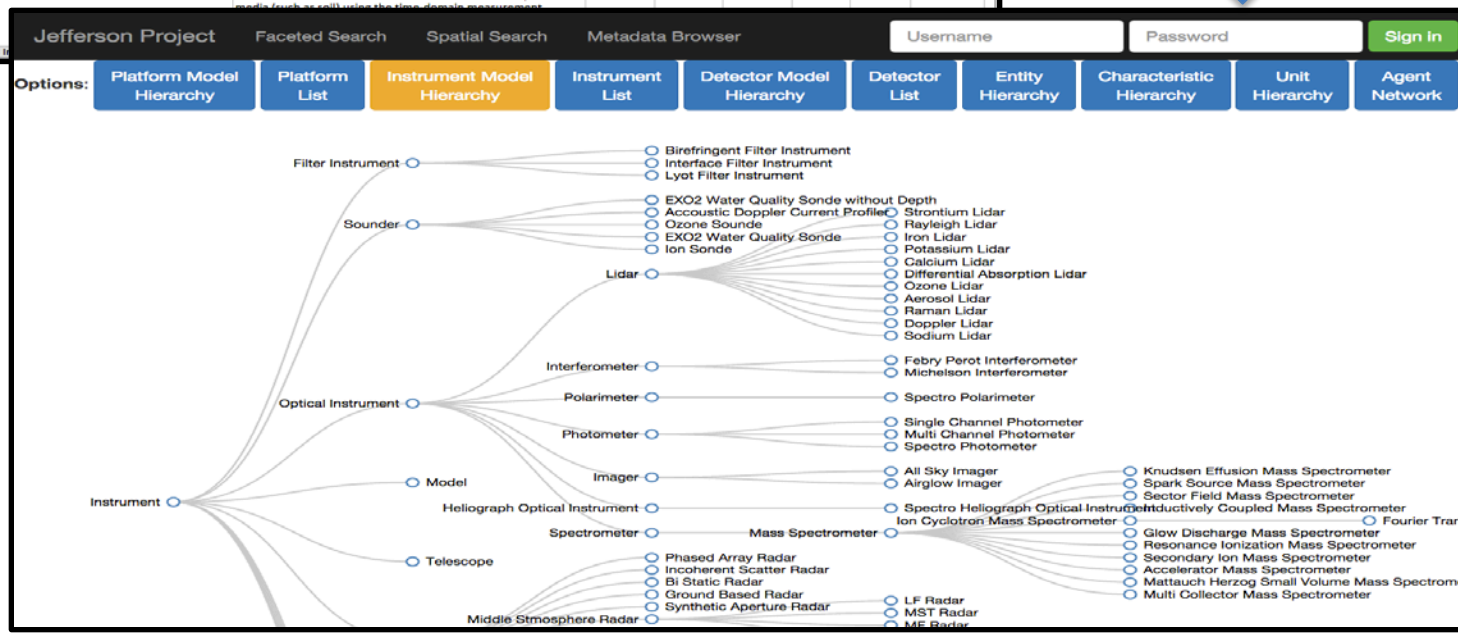
Metadata Browser: From Spreadsheets to Web Visualizations

Spreadsheets have been shared between IBM, RPI and the FUND for Lake George



	rdfs:label	vstoi:hasMaker	rdfs:comment	hasImage	vstoi:numAt	vstoi:maxDe	vstoi:maxl	vstoi:minO	vstoi:maxOpe
1	Accoustic Doppler Current Profiler								
2	Weather Station								
4	EXO2 Water Quality Sonde	http://jefferson.tw.rpi.edu/ontology/jp-kb.owl#Org-YSI	The EXO2 water quality sonde is a YSI multiparameter instrument that collects data with six user-replaceable sensors, a central wiper to keep sensors clean of biofouling, and an integral pressure transducer for depth.		1	6	1000000	5	50
5	EXO2 Water Quality Sonde without Depth	http://jefferson.tw.rpi.edu/ontology/jp-kb.owl#Org-YSI	The EXO2 water quality sonde is a YSI multiparameter instrument that collects data with six user-replaceable sensors, a central wiper to keep sensors clean of biofouling, and an integral pressure transducer for depth.		0	6	1000000	5	50
6	SONTEK-IQ	http://jefferson.tw.rpi.edu/ontology/jp-kb.owl#Org-Xylen	Instrument used for monitoring flows in canals, culverts, pipes, and natural streams, which includes a custom flow algorithm derived from hundreds of field measurements. Four velocity beams profile water velocity in 3-D - both vertically and horizontally - ensuring complete coverage of the velocity field. The built-in pressure sensor and vertical acoustic beam work in tandem to measure water level.		5	0			
7	Vaisala WXT520	http://jefferson.tw.rpi.edu/ontology/jp-sn.owl#Org-Vaisala	The Vaisala Weather Transmitter WXT520 measures barometric pressure, humidity, precipitation, temperature, and wind speed and direction.To measure wind speed and direction, the WXT520 has the Vaisala WINDCAP® Sensor that uses ultrasound to determine horizontal wind speed and direction. The array of three equally spaced transducers on a horizontal plane is a Vaisala specific design. Barometric pressure, temperature, and humidity measurements are combined in the PTU module using capacitive measurement for each parameter. It is easy to change the module without any contact with the sensors.		5	0		-52	60
8	BP sensor PTB 110	http://jefferson.tw.rpi.edu/ontology/jp-sn.owl#Org-Vaisala	The Vaisala BAROCAP® Barometer PTB110 is designed both for accurate barometric pressure measurements at a room temperature and for general environmental pressure monitoring over a wide temperature range.		1	0	0	-40	60
9	HC253 Temperature and Relative Humidity Prot	http://jefferson.tw.rpi.edu/ontology/jp-sn.owl#Org-CampbellScientific	The HC253 is a rugged, accurate temperature/RH probe used for longterm, unattended applications. The probe uses a Rotronic's IN1 capacitive sensor to measure RH and a 100 ohm PRT to measure temperature.		2	0		-40	100
10	LiCor LI200x pyranometer	http://jefferson.tw.rpi.edu/ontology/jp-sn.owl#Org-LiCor	The LI-200R Pyranometer measures total solar radiation.		1	0	0		
11	TB3 Tipping Bucket Rain Gauge	http://jefferson.tw.rpi.edu/ontology/jp-sn.owl#Org-HydrologicalServices	TB3 Tipping Bucket Rain Gauge is used for measuring rainfall and precipitation in remote and unattended locations		1	0	0	-20	70
12	Campbell Model 107 Temperature Probe	http://jefferson.tw.rpi.edu/ontology/jp-sn.owl#Org-CampbellScientific	The 107 Temperature Probe uses a thermistor to measure temperature in air, soil, and water.		1	0		-50	100
			The CS616 measures the volumetric water content of porous media (such as soil) using the time domain measurement						

PlatformModels | Platforms | InstrumentModels

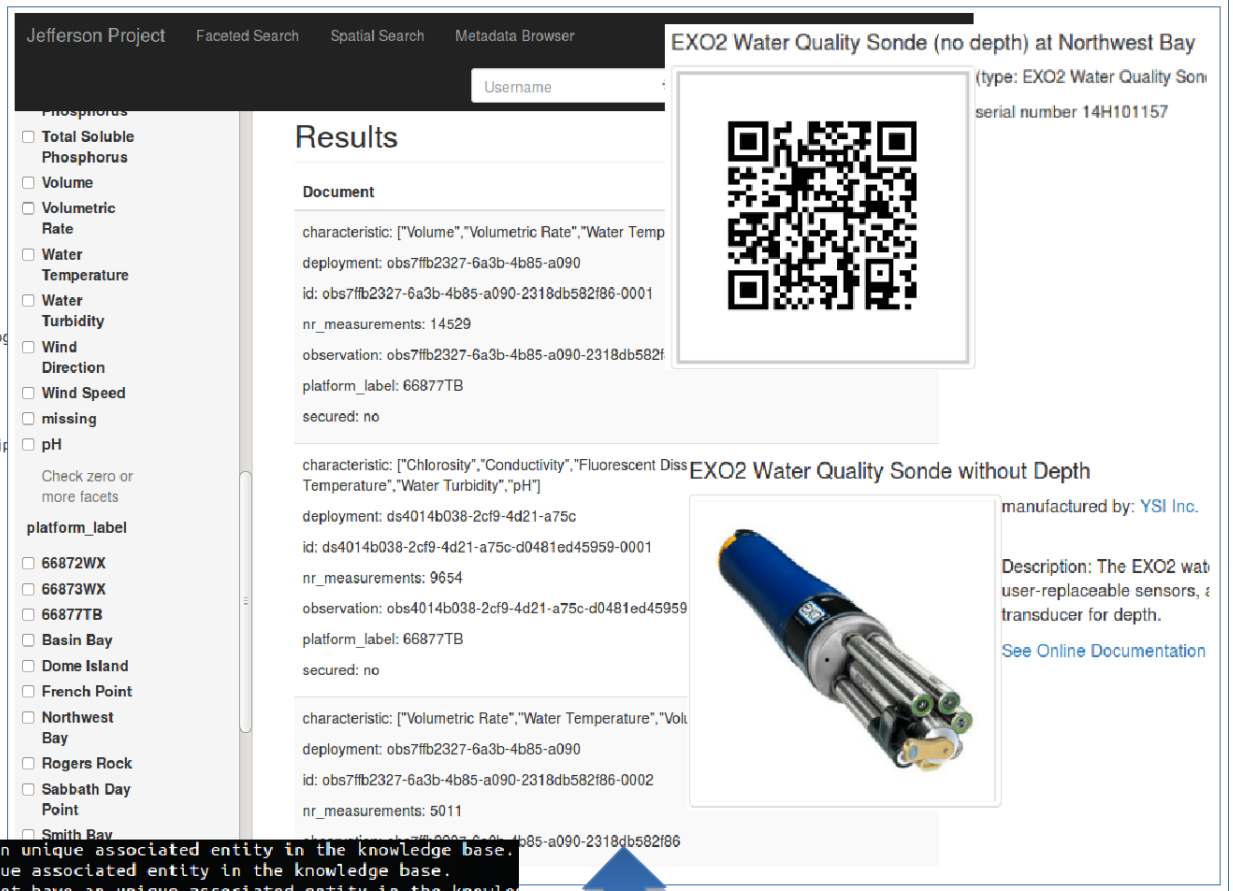


Spreadsheet content is now stored in a RDF knowledge repository based on HASNetO vocabulary

Data Collection and Dissemination

```
<jp-kb>
        a ccsv:KnowledgeBase;
        ccsv:hasConnectionURL "http://jefferson.tw.rpi.edu/ontolog

<deployment-TEST-001>
        a vstoi:Deployment;
        prov:startedAtTime "2014-02-10T01:01:01Z"^^xsd:dateTime;
        vstoi:hasInstrument <instrument-TEST-001>;
        vstoi:hasPlatform <http://jefferson.tw.rpi.edu/ontology/jp

<observation-TEST-001>
        a oboe:Observation;
        a time:Interval;
        prov:startedAtTime "2014-02-10T01:02:01Z"^^xsd:dateTime .

<deployment-TEST-001>
        hasneto:hasObservation <observation-TEST-001> .

<dataset-TEST-001>
        a vstoi:Dataset;
        prov:wasGeneratedBy <observation-TEST-001>;
        hasneto:hasMeasurementType <mt0> .

<mt0>
        a oboe:Measurement;
        a time:Instant;
        time:inDateTime <ts0>;
        ccsv:atColumn 2;
        oboe:ofCharacteristic jp-characteristics:WindSpeed;
        oboe:usesStandard oboe-standards:MeterPerSecond .

<ts0
```

CCSV datasets




```
[ERROR] Characteristic TotalNitrogen does not have an unique associated entity in the knowledge base.
[ERROR] Characteristic Calcium does not have an unique associated entity in the knowledge base.
[ERROR] Characteristic TotalSolublePhosphorus does not have an unique associated entity in the knowle
[ERROR] Characteristic Nitrate does not have an unique associated entity in the knowledge base.
[ERROR] Characteristic Potassium does not have an unique associated entity in the knowledge base.
[ERROR] Characteristic WaterTemperature does not have an unique associated entity in the knowledge bas
[ERROR] Characteristic Sodium does not have an unique associated entity in the knowledge base.
[ERROR] Characteristic Ammonium does not have an unique associated entity in the knowledge base.
[ERROR] Characteristic Alkalinity does not have an unique associated entity in the knowledge base.
```

1. validate

HADataC-Loader

2. store

SOLR-HADataC

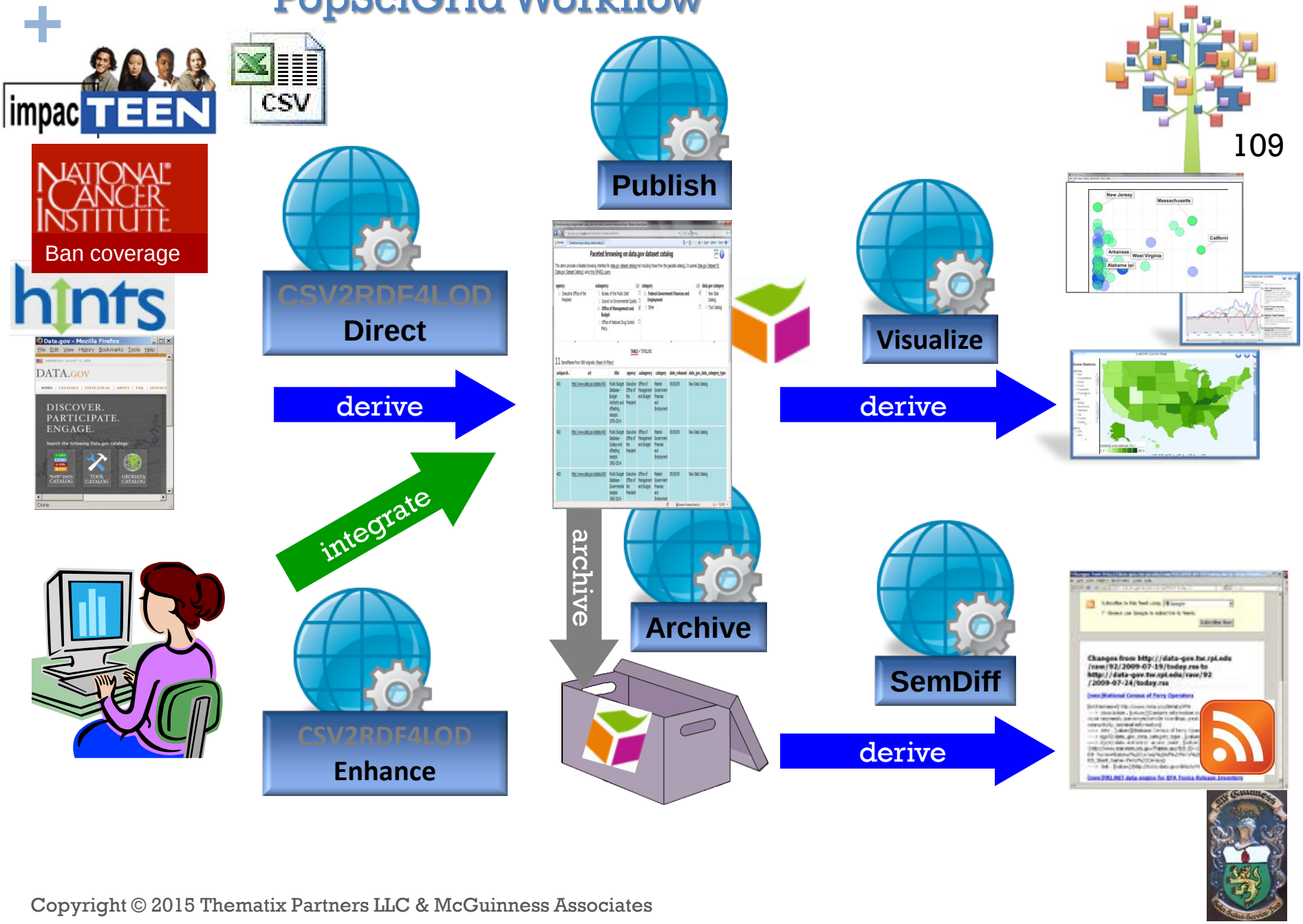
Measurements | Metadata

Population Sciences Grid Goals

- Convey complex health-related information to consumer and public health decision makers for community health impact leveraging the growing evidence base for communicating health information on the Internet
- Inform the development of future research opportunities effectively utilizing cyberinfrastructure for cancer prevention and control
- How can semantic technologies be used to integrate, present, and analyze data for a wide range of users?
- Can tools allow lay people to build their own demos and support public usage and accurate interpretation?
- Within PopSciGrid:
  - Which policies (taxation, smoking bans, etc) impact health and health care costs?
  - What data should be displayed to help scientists and lay people evaluate related questions?
  - What data might be presented so that people choose to make (positive) behavior changes?
  - What does the data show? why should someone believe that?
  - What are appropriate follow ups?

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

PopSciGrid Workflow



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

+ Multi-dimensional Semantic Integration and Analysis (Smoking)

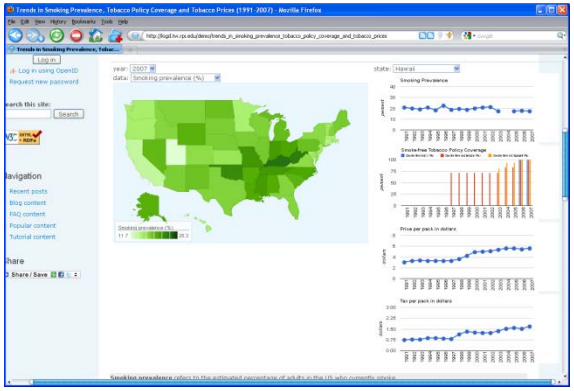
Ex. Question:

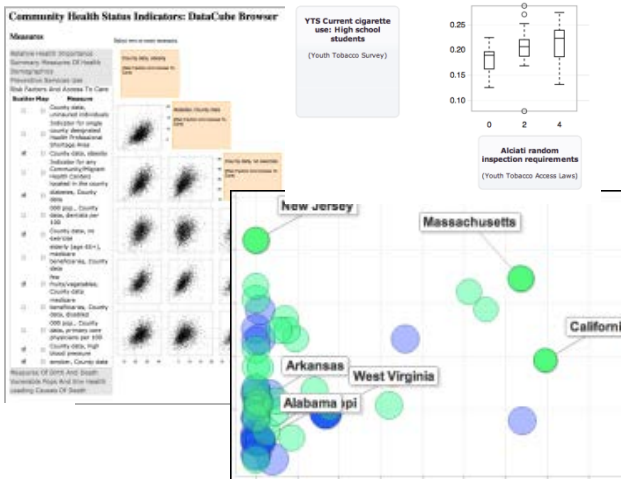
- What intervention strategies correlate with decreased smoking? - By age group, geographic area, etc.

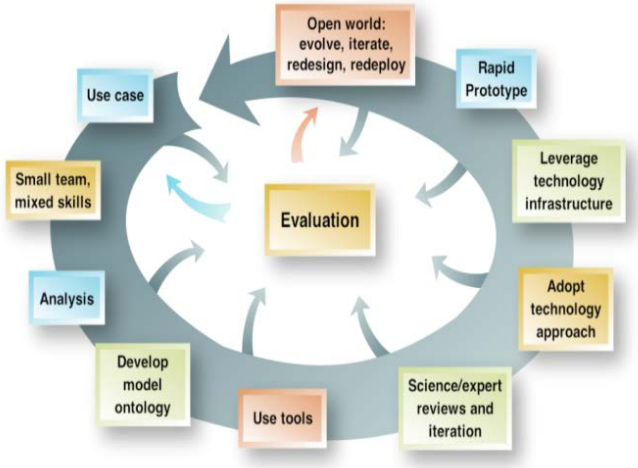
Semantic Web methodology, accountable mashups, multi-dimensional analysis, aggregation semantics tools

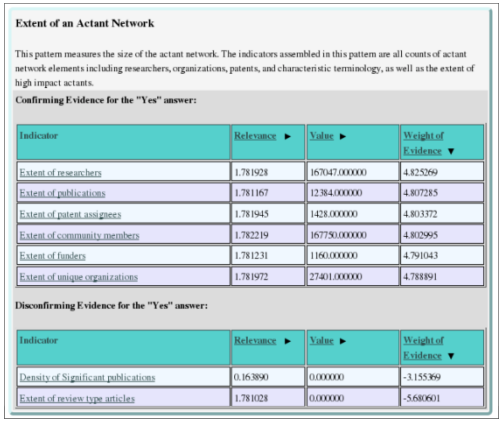
Domain answer: bans and labeling for certain age bracket

PopSciGrid with NIH looked at "preventable cancers" , hypothesized contributors (smoking), and interventions – taxation, bans, labeling









Extent of an Actant Network

This pattern measures the size of the actant network. The indicators assembled in this pattern are all counts of actant network elements including researchers, organizations, patents, and characteristic terminology, as well as the extent of high impact actants.

Confirming Evidence for the "Yes" answer:

Indicator	Relevance ►	Value ►	Weight of Evidence ▼
Extent of researchers	1.781928	167047.000000	4.825269
Extent of publications	1.781167	12384.000000	4.807285
Extent of patent assignees	1.781945	1428.000000	4.803372
Extent of community members	1.782219	167750.000000	4.802995
Extent of funders	1.781231	1160.000000	4.791043
Extent of unique organizations	1.781972	27401.000000	4.788891

Disconfirming Evidence for the "Yes" answer:

Indicator	Relevance ►	Value ►	Weight of Evidence ▼
Density of Significant publications	0.163890	0.000000	-3.155369
Extent of review type articles	1.781028	0.000000	-5.680601

Multi-dimensional Semantic Analysis work also being developed for the IARPA FUSE (Foresight and Understanding from Scientific Exposition Project

McCusker,McGuinness,Lee,Thomas,Courtney,Tatalovich,Contractor,Morgan,Shaikh. Towards Next Generation Health Data Exploration: A Data Cube-based Investigation into Population Statistics for Tobacco, 2013

Michaelis, J., McGuinness, D.L., Chang, C., Hunter, D., and Babko-Malaya, O. 2013. Towards Explanation of Scientific and Technological Emergence. Proc. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining Niagara Falls, Canada.

Reflections

- Uses semantic technologies along with simple ontologies and provenance-encoding tools
- Provides an operational specification for other policy exploration tools – initially tobacco data and related policies, later physical education and nutrition policies in elementary, junior and senior high school (CLASS - http://class.cancer.gov/About.aspx) and laid foundation for broader multidimensional analysis projects
- Similar model being used in the Tetherless SemantEco / SemantAqua / SemantAir Portal family

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Why did I present wine, water, and linked data applications?

- Wine advisor shows semantic technology in action making actionable recommendations and mobile health advisor shows explanations of data as well as linked data
- Water ecosystems applications show a "typical" semantic integration web 3.0 application as well as a migration to broader complex
- Both of these styles are needed for many semantic applications and these features are realizable today
- Next – they depend on
  - Understandable and Actionable applications
  - Semantic web stack
  - Data availability – linked data cloud is growing
  - Ontologies
  - Semantic methodologies

TetherlessWorld
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Foundations: Trust & Understanding

If users (humans and agents) are to use, reuse, and integrate system answers, they must trust them.

System transparency supports understanding and trust.

Even simple "lookup" systems benefit from providing information about their sources.

Systems that manipulate information (with sound deduction or potentially unsound heuristics) benefit from providing information about their manipulations.

Goal: Provide interoperable infrastructure that supports explanations of sources, assumptions, and answers as an enabler for trust.

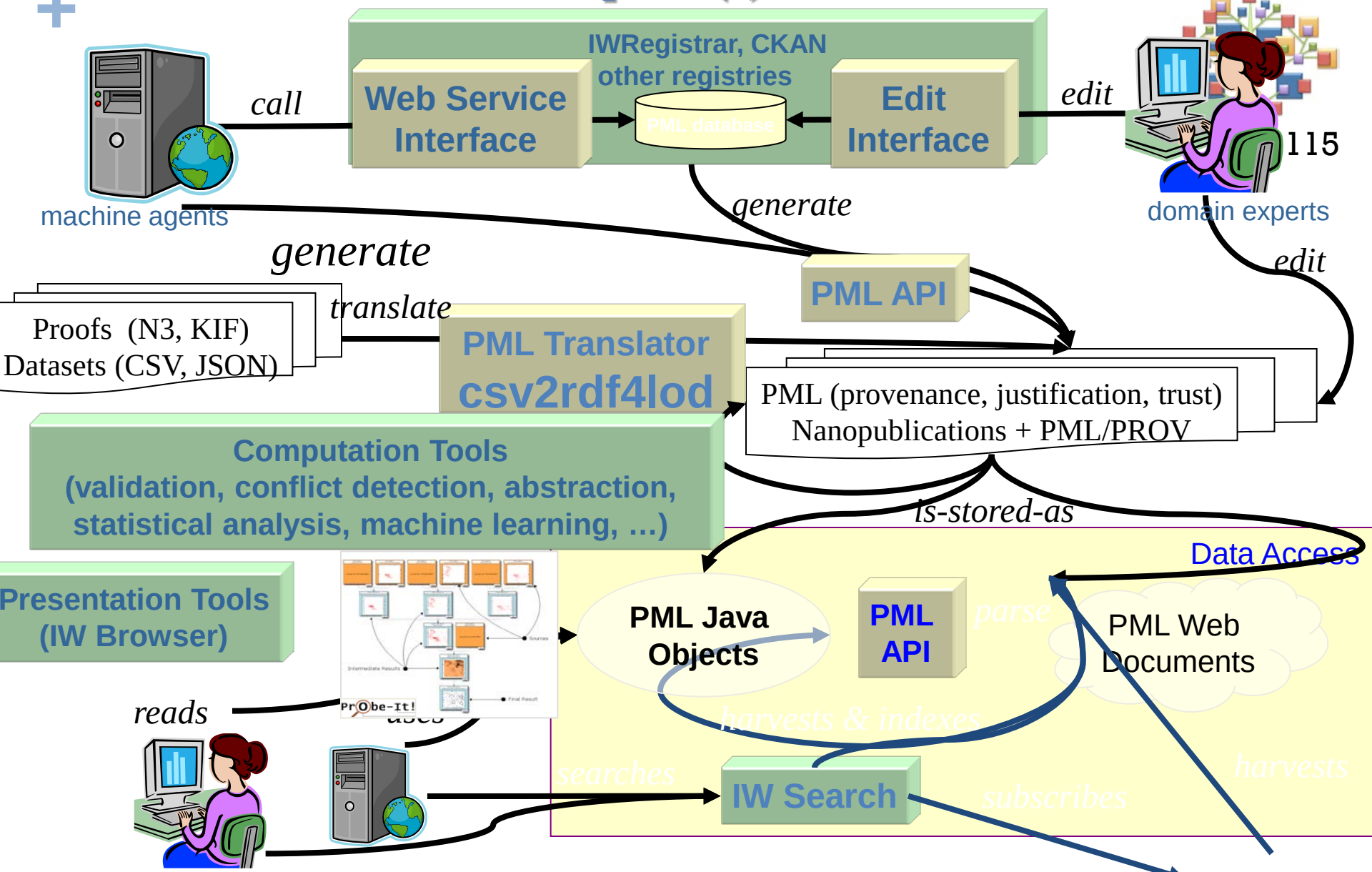
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Foundations: Explanations, Proof Analysis Inference Web 1.0

- Framework for explaining question answering tasks
  - Stores and manages meta-information about proofs and explanations through a distributed repository
  - Designed to use a declarative Provenance Markup Language (originally used the OWL-based Proof Markup Language (PML) for proof interchange. PML contributed to W3C's PROV
- Services include proof presentation, strategy-based views, filtering, trust, combination, expansion, checking, search, and other capabilities
- Integrated browsing and display of provenance documents from diverse sources
- Rewriting capabilities for improved understanding
- Multi-modal dialogue options including alternative strategies for presenting explanations, visualizations, and summaries

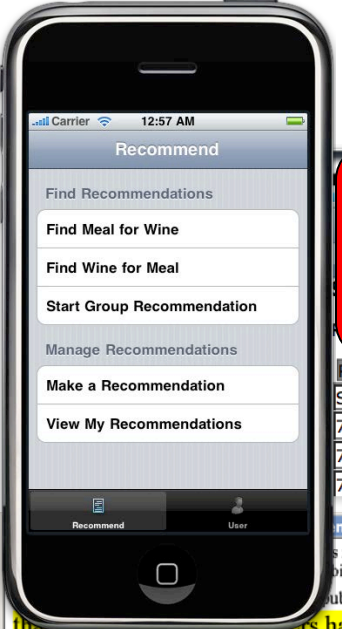
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Technical Architecture Option(s)

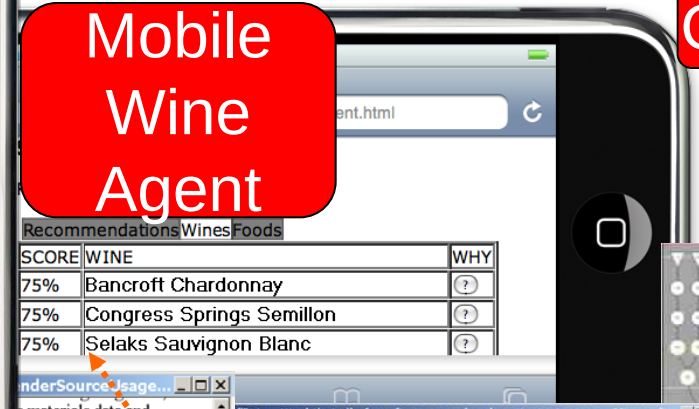


Copyright © 2015 Thematix Partners LLC & McGuinness Associates

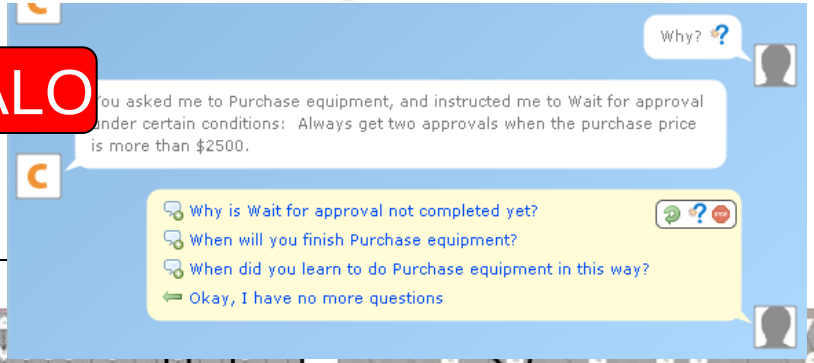
Making Systems Actionable with Knowledge Provenance

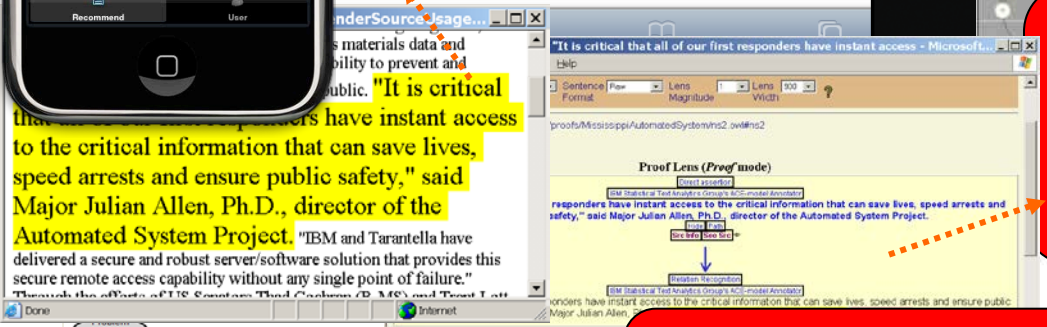


Mobile Wine Agent

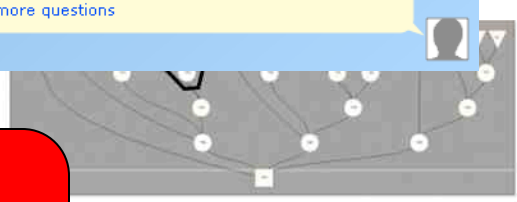


CALO

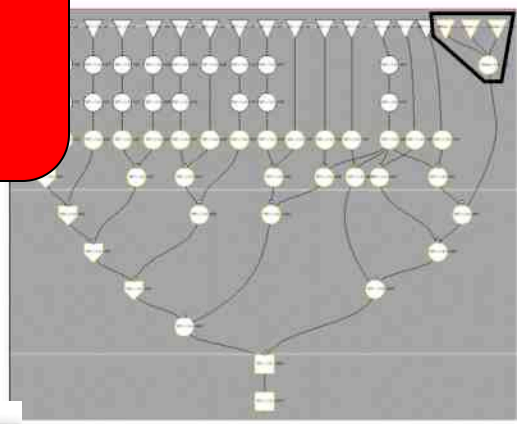




Combining Proofs in TPTP

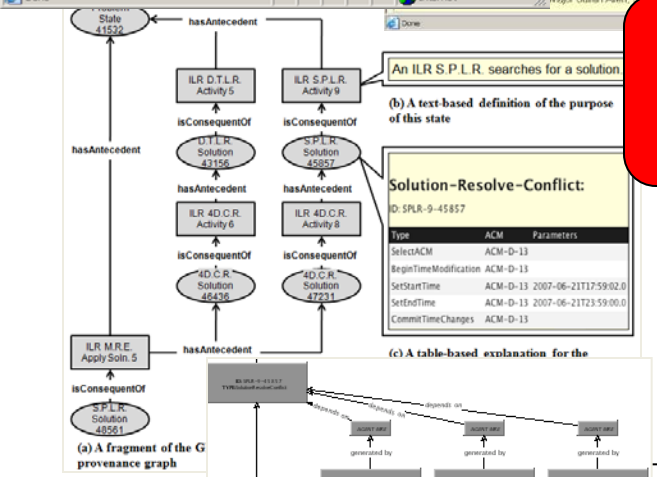


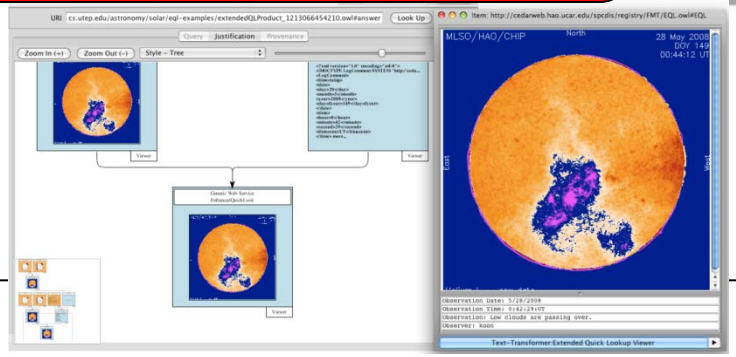
Otter's proof, helper subproof outlined



Combined proof, replaced subproof outlined

Intelligence Analyst Tools





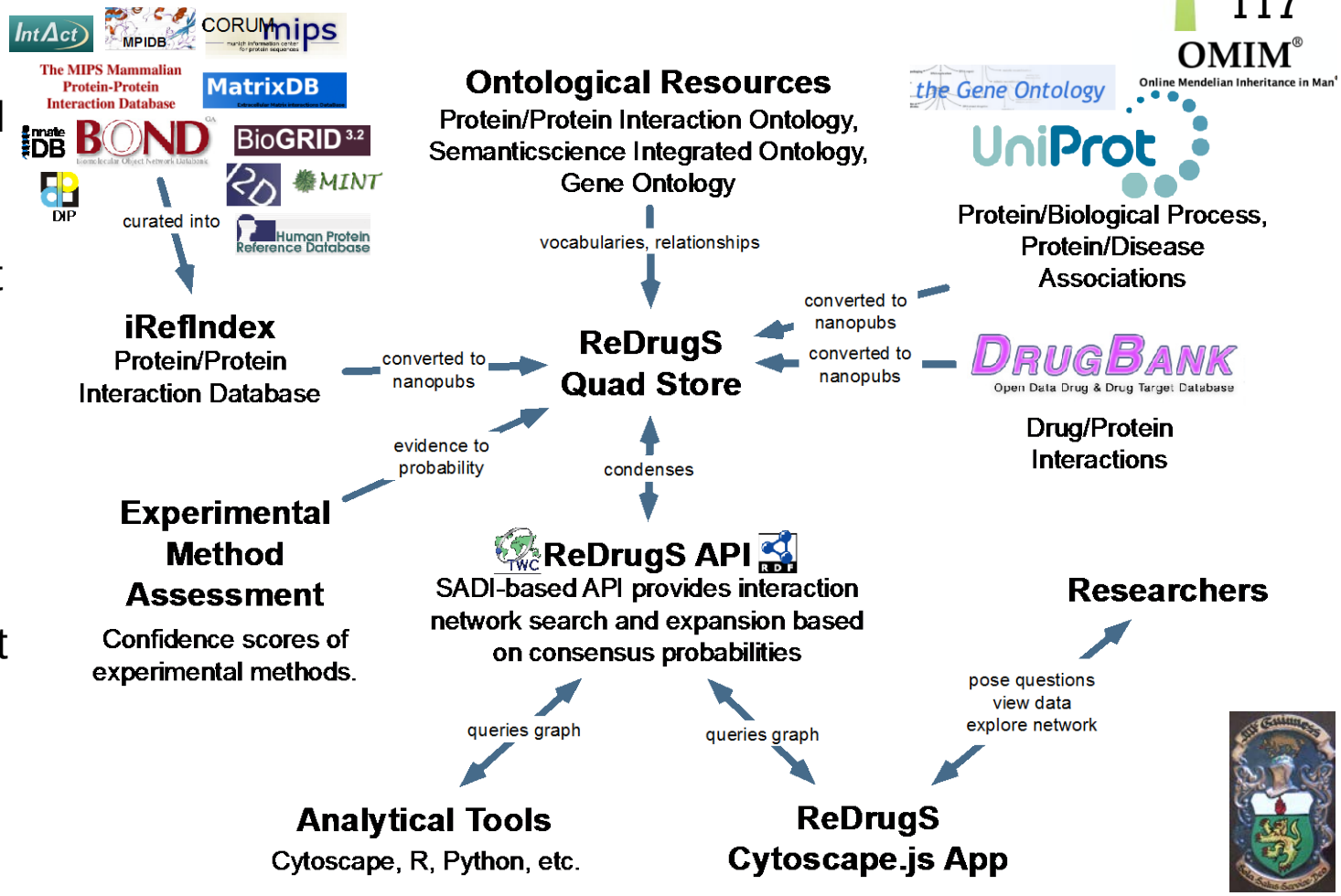
Knowledge Provenance in Virtual Observatories

GILA

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

ReDrugS
(Repurposing of Drugs using Semantics)

- Use semantic technologies to encode and process biological knowledge to generate hypotheses about new uses for existing drugs.
- Leverage existing curated data sources, build reusable integrated content sources and infrastructure

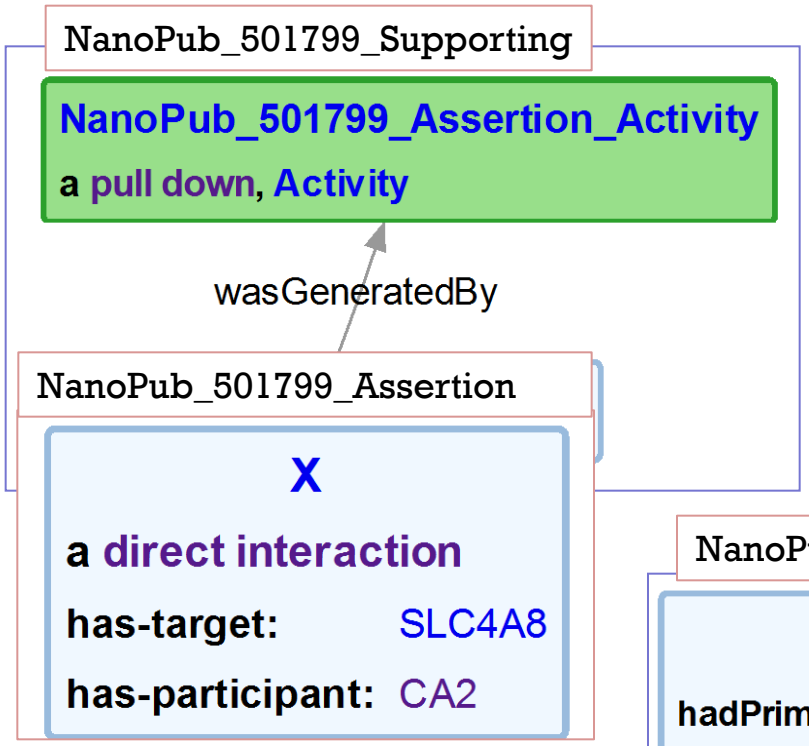


McCusker, J., Solanki, K., Chang, C., Dumontier, M., Dordick, J., and McGuinness, D.L. 2014. A Nanopublication Framework for Systems Biology and Drug Repurposing. Proc. of CSHALS 2014 Boston, MA.

McCusker, J., Yan, R., Solanki, K., Erickson, J.S., Chang, C., Dumontier, M., Dordick, J., and McGuinness, D.L. 2014. A Nanopublication Framework for Biological Networks using Cytoscape.js. In Proceedings of International Conference on Biomedical Ontologies (ICBO 2014) (October 6-9 2014, Houston, TX).

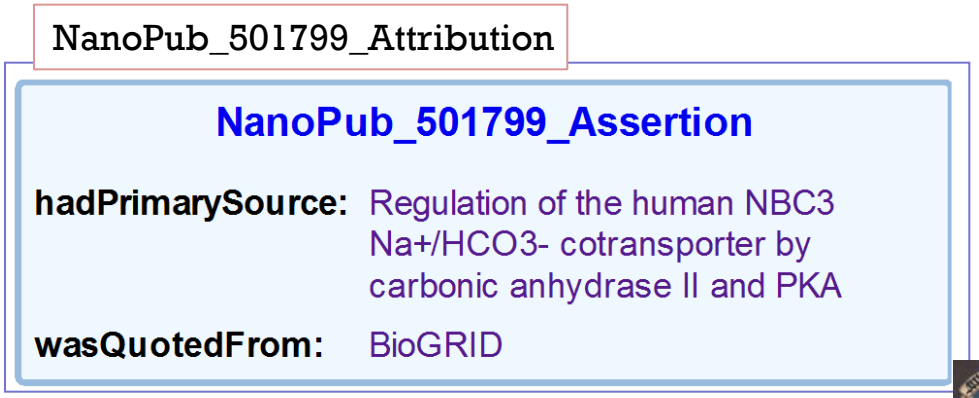
http://tw.rpi.edu/web/doc/redrugsnanopub

Nanopublications



Tetherless World Constellation RPI

Simple yet semantically-rich encodings allow algorithms to not just find correlation but to look for causality using reasoning



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Experimental Method Coverage

- 99.98% coverage of the ~936,000 nanopubs with evidence data from iRefIndex.

- Top 10 methods (86% coverage):

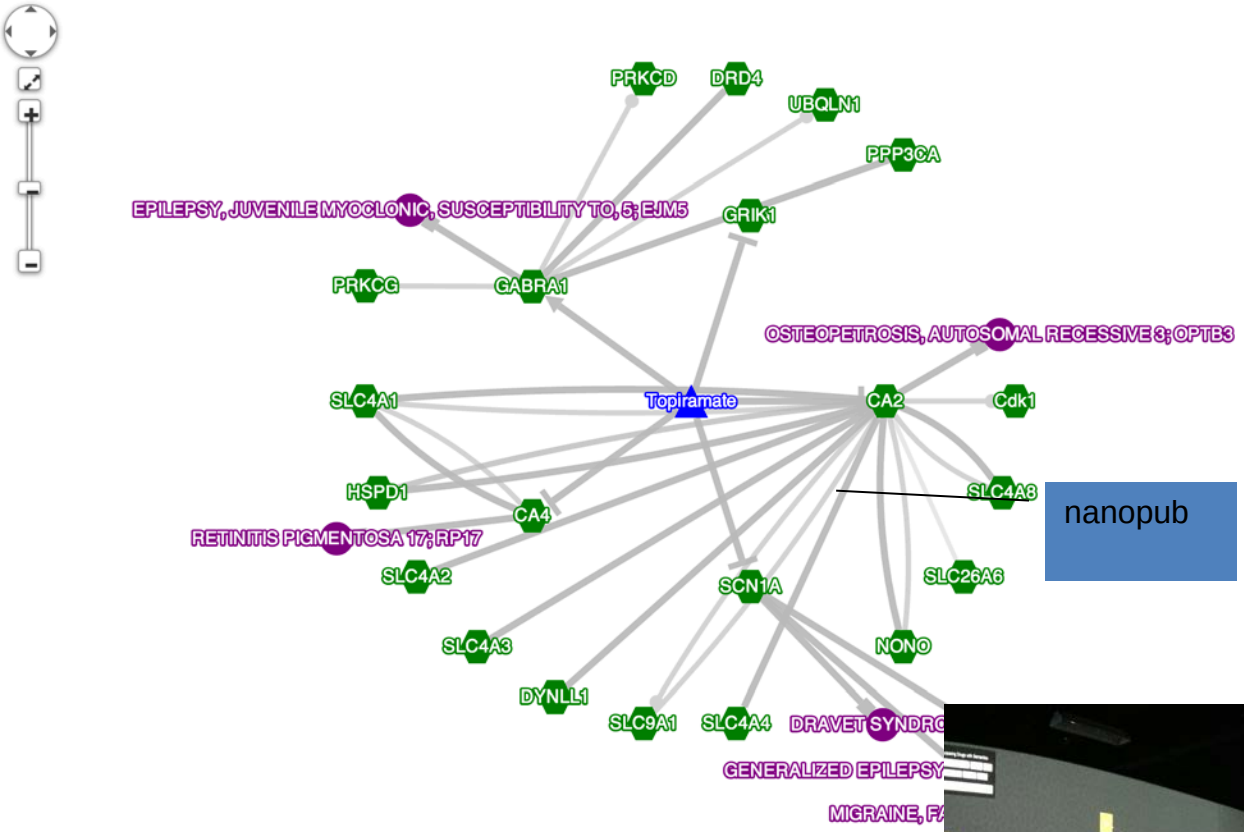
Method	Count	P conf	M conf
two hybrid [mi:0018]	199130	1	1
genetic interference [mi:0254]	196717	2	2
affinity chromatography technology [mi:0004]	117659	2	2
tandem affinity purification [mi:0676]	70545	2	2
two hybrid pooling approach [mi:0398]	60715	1	1
anti tag coimmunoprecipitation [mi:0007]	42249	3	3
pull down [mi:0096]	37676	2	2
two hybrid array [mi:0397]	29806	1	1
x-ray crystallography [mi:0114]	29182	2	3
anti bait coimmunoprecipitation [mi:0006]	22533	2	3

Tetherless World Constellation, RPI

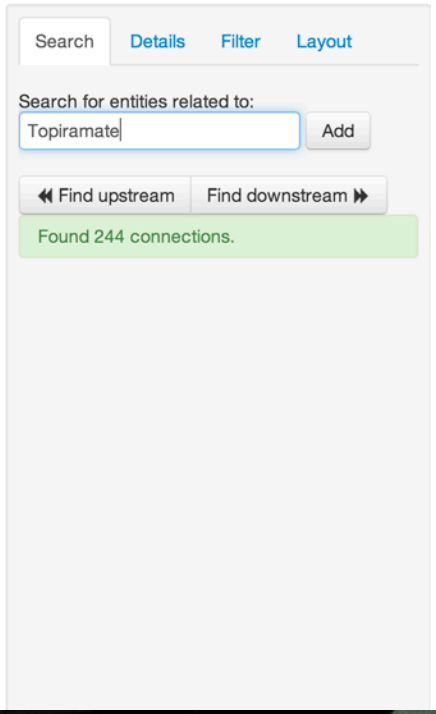
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

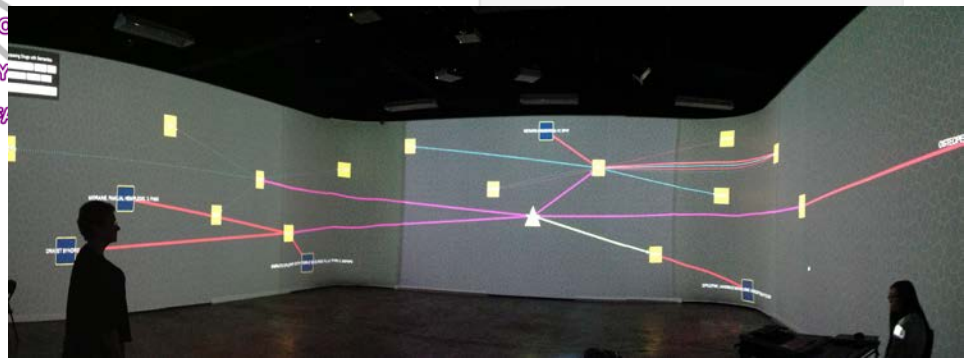
Topiramate Disease Associations: p ≥ 0.9

Repurposing Drugs with Semantics



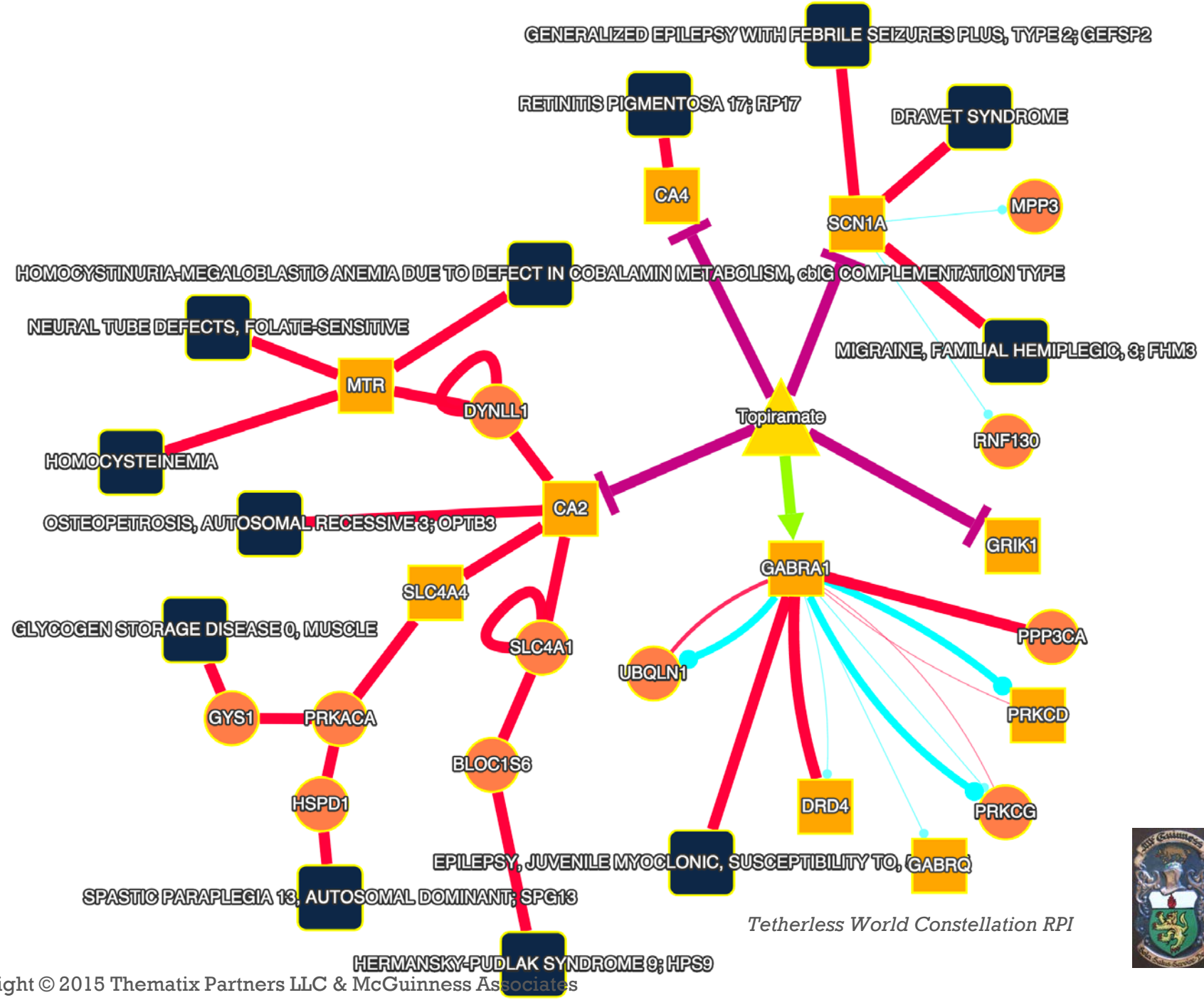
nanopub





Tetherless World Constellation RPI

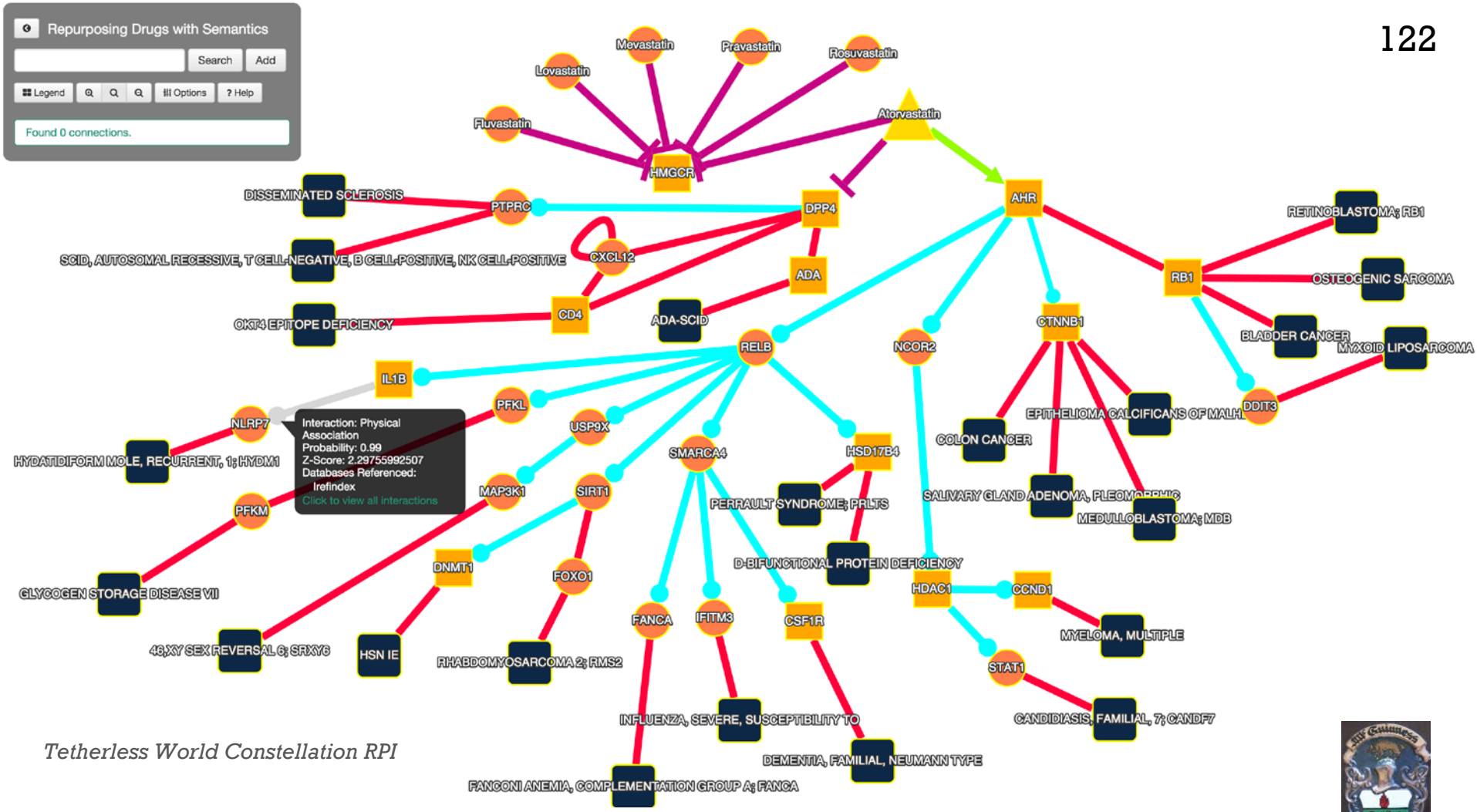
Copyright © 2015 Thematix Partners LLC & McGuinness Associates



Tetherless World Constellation RPI

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Atorvastatin (Lipitor)



Tetherless World Constellation RPI

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

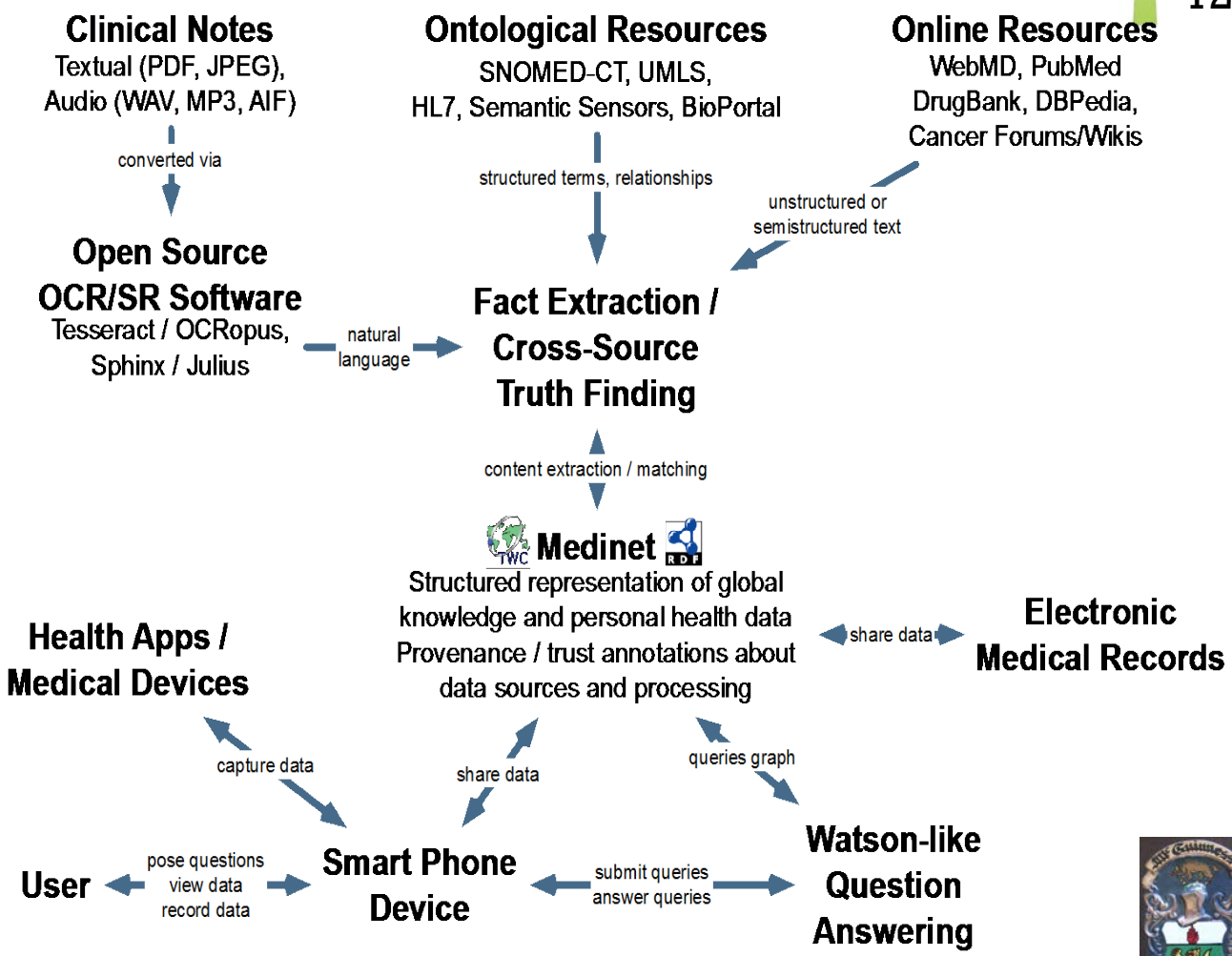
Reflections

- We are just scratching the surface here but this holds potential to support more exploration related to causation rather than just correlation.
- This infrastructure is also used in SemNext – Semantic Numeric Exploration Technologies – currently on brain data.
- Looking at reuse in a microbiome project
- Broad reusage.
- Simpler provenance model based on lightweight use of PROV with nanopublications

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

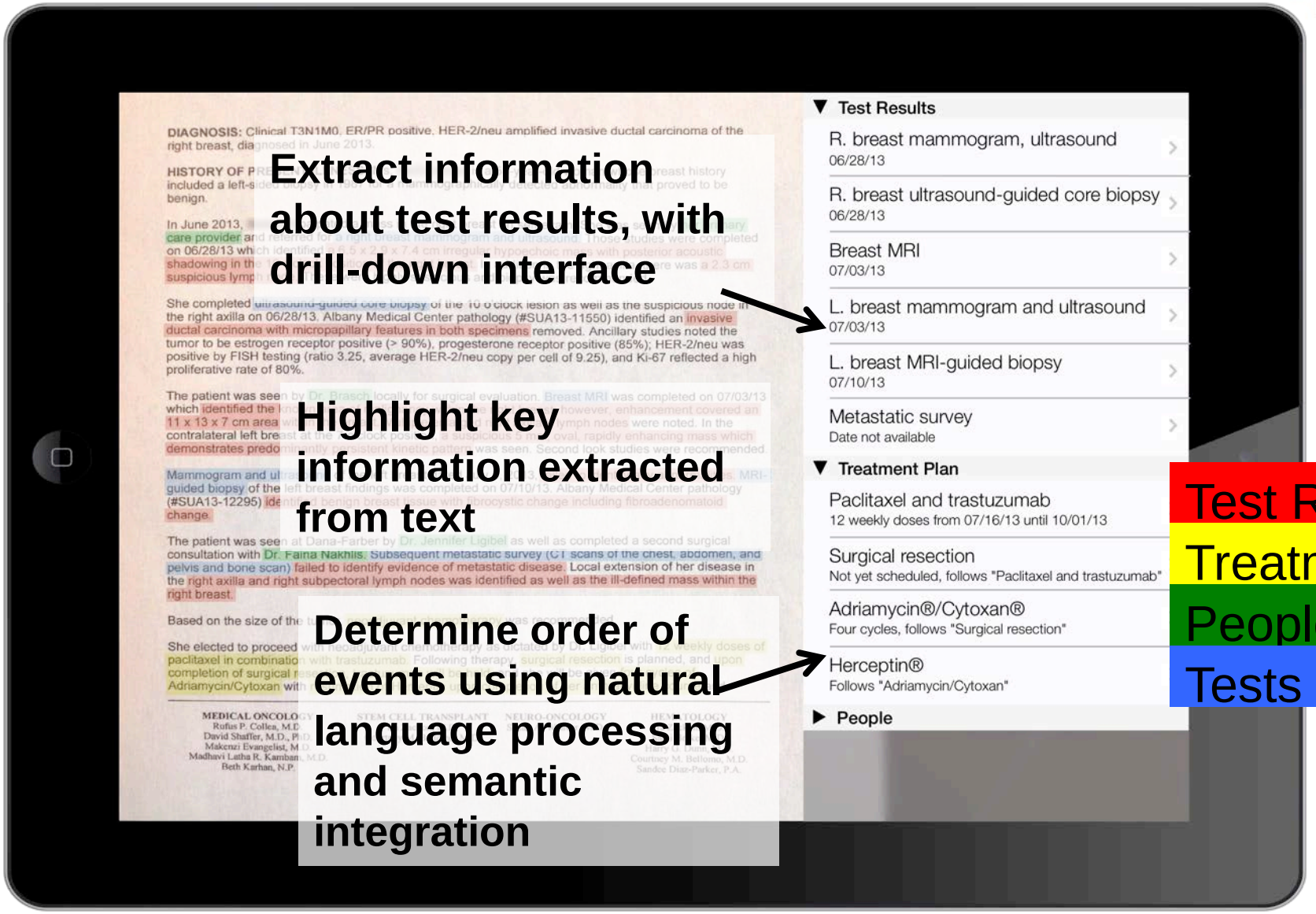
Mobile Health Assistant

- Use semantic technologies to encode, and integrate a wide range of health information to help people function at a level higher than their training (patients become more educated, docs find more relevant content,..).
- Leverage existing curated **and** uncurated sources, build reusable integrated content sources and infrastructure



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Doctor's Perspective



Extract information about test results, with drill-down interface

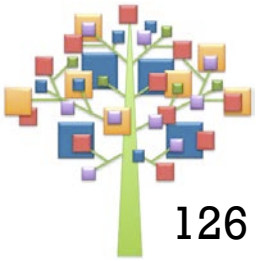
Highlight key information extracted from text

Determine order of events using natural language processing and semantic integration

Test Results
Treatments
People
Tests

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Patient's Perspective



Link to external resources for descriptive information

Identify possible side-effects and coping strategies

Focus on information the patient is concerned about

Explanation of reasoning and why information may be unavailable

Test Results
Treatments
People
Tests

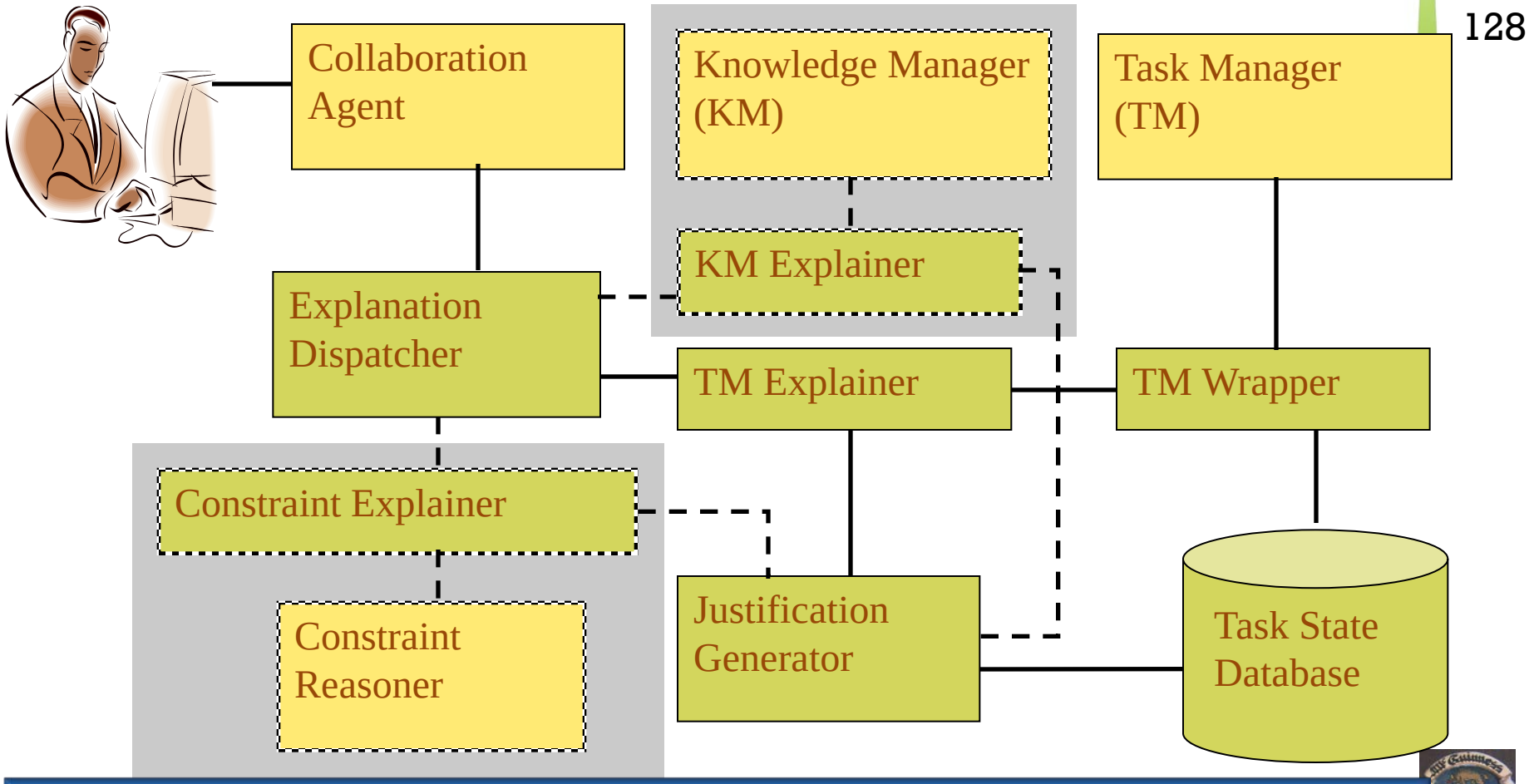
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Cognitive Assistant that Learns & Organizes

- DARPA IPTO funded program
- Personal office assistant, tasked with:
  - Noticing things in the cyber and physical environments
  - Aggregating what it notices, thinks, and does
  - Executing, adding/deleting, suspending/resuming tasks
  - Planning to achieve abstract objectives
  - Anticipating things it may be called upon to do or respond to
  - Interacting with the user
  - Adapting its behavior in response to past experience, user guidance
- 22 participating organizations
- Led to other efforts including Siri (later acquired by Apple and in iPhone

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

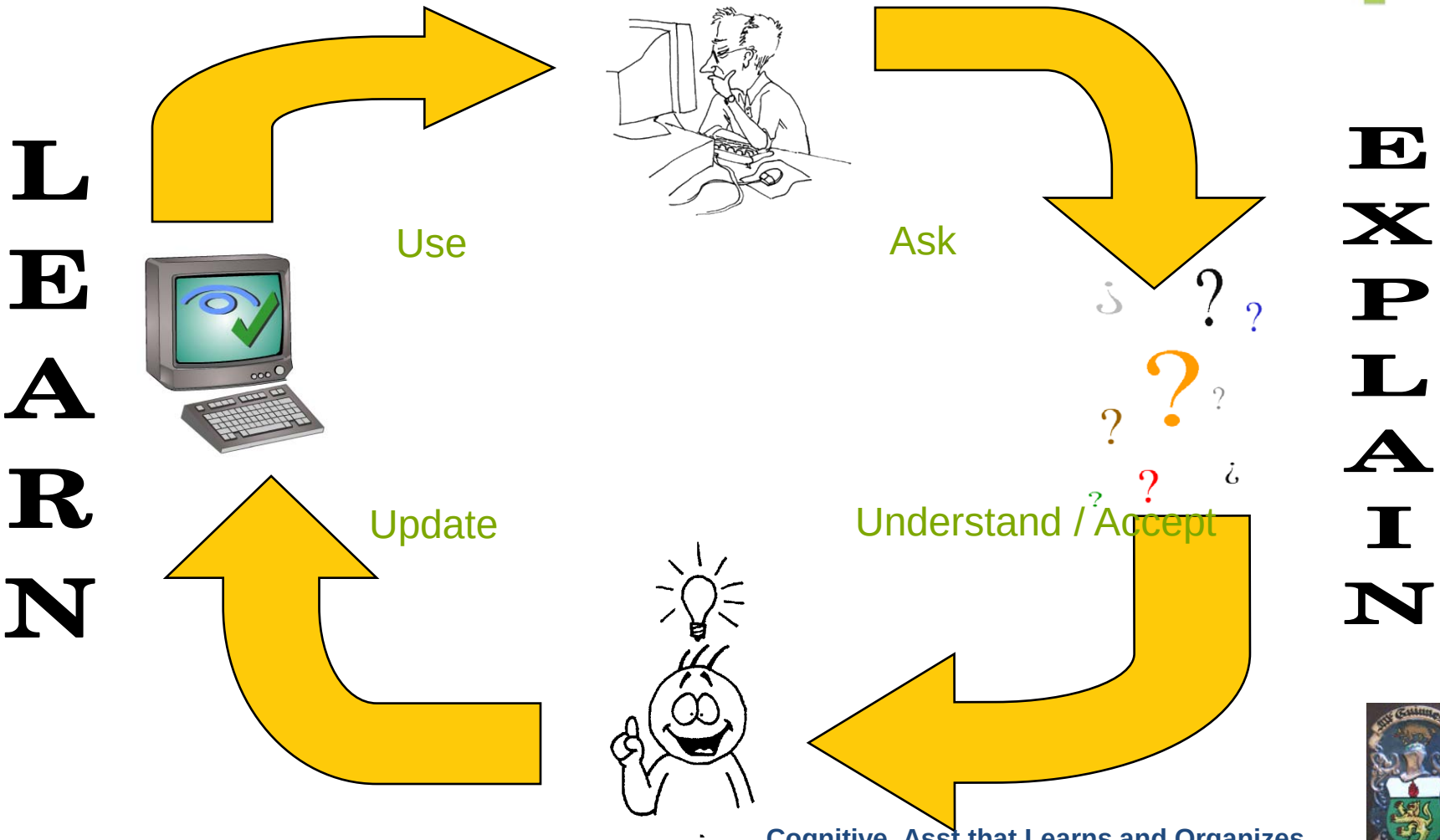
Integrated Cognitive Explanation Environment for Cognitive Asst that Learns and Organizes



A unified framework for explaining behavior and reasoning is essential for users to trust and adopt cognitive assistants.

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

The Use-Ask-Understand-Update Cycle



Cognitive Asst that Learns and Organizes (CALO) explanation is joint work with McGuinness, Glass, Wolverton, Chang, Ding

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Summary

- Introduced Semantic Technology Foundations
- Provided numerous examples of functioning systems with value propositions, many users, etc.
- Foundational technology is available from many sources

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Elisa Kendall
Partner, Thematix Partners LLC
(408) 930-5601
ekendall @ thematix.com
http://thematix.com

Questions?

Deborah McGuinness
Tetherless World Constellation Chair,
Professor of Computer Science and Cognitive Science
Rensselaer Polytechnic Institute and
CEO, McGuinness Associates
(518) 276-4404
dlm @ cs.rpi.edu
http://tw.rpi.edu/web/person/Deborah_L_McGuinness

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Acronym Soup

- CL – ISO 24707 Common Logic: a family of first order logic languages, including Conceptual Graphs & Common Logic Interchange Format – a successor to the Knowledge Interchange Format (KIF), http://cl.tamu.edu/
- DAML – DARPA Agent Mark-up Language, one of the primary languages leading to the development of OWL, http://www.daml.org/
- DARPA – Defense Advanced Research Projects Agency, http://www.darpa.mil/
- DL – Description Logics: a subset of first order logic, for which tractable & complete reasoning systems are available
- IRI – Internationalized Resource Identifier
- Linked Data and RDFa, http://www.w3.org/standards/techs/rdfa#w3c_all
- MDA – Model-Driven Architecture, http://www.omg.org/mda/
- ODM – Ontology Definition Metamodel, http://www.omg.org/spec/ODM/1.0/
- OWL – W3C Web Ontology Language, OWL 2 specifications dated 27 October 2009, http://www.w3.org/standards/techs/owl#w3c_all
- OWL-S – a set of OWL ontology components that extend the W3C OWL specifications to support Semantic Web Services, http://www.daml.org/services/owl-s/1.1/

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

More Acronym Soup

- PML – Proof Markup Language – Provenance Interlingua, www.inference-web.org
- PRR – Production Rules Representation, http://www.omg.org/spec/PRR/1.0/
- RIF – Rule Interchange Format, http://www.w3.org/standards/techs/rif#w3c_all
- RDF – Resource Description Framework, http://www.w3.org/TR/rdf-concepts/
- Semantic Annotations for WSDL and XML (SAWSDL), http://www.w3.org/standards/techs/sawsdl#w3c_all
- Simple Knowledge Organization System (SKOS), http://www.w3.org/standards/techs/skos#w3c_all
- SPARQL – Semantic Web Query Language, http://www.w3.org/standards/techs/sparql#w3c_all
- Semantic Web Rule Language (SWRL), http://www.w3.org/Submission/SWRL/
- URI – Uniform Resource Identifier
- WSDL – Web Services Description Language

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Foundations: Web Layer Cake

Visualization APIs
S2S
Govt Data

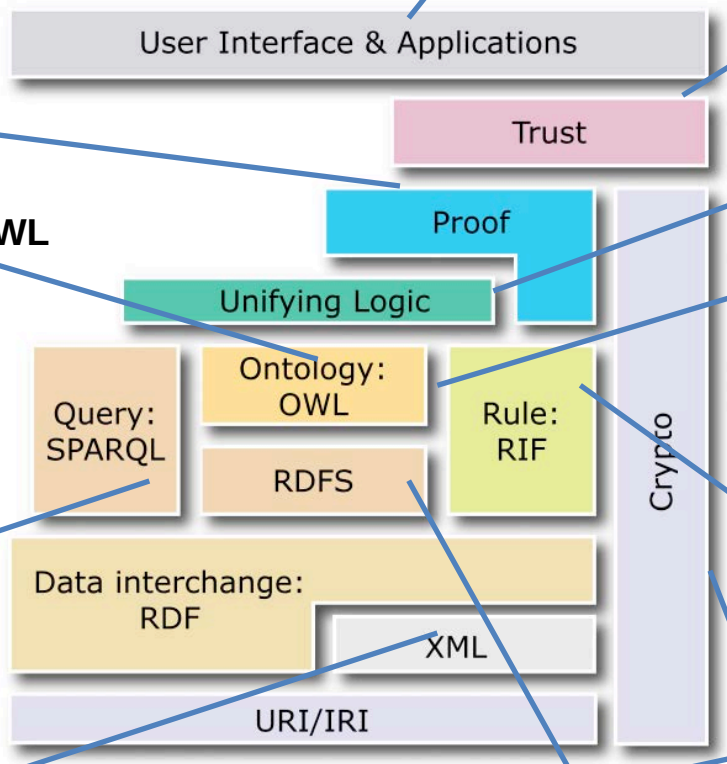
Inference Web, Proof Markup Language, W3C Provenance Working group formal model, W3C incubator group, …

Inference Web IW Trust, Air + Trust

DL, KIF, CL, N3Logic

OWL 1 & 2 WG Edited main OWL Docs, quick reference, OWL profiles (OWL RL), Earlier languages: DAML, DAML+OIL, Classic

Ontology repositories (ontolinguag), Ontology Evolution env: Chimaera, Semantic eScience Ontologies, MANY other ontologies



SPARQL WG, earlier QL – OWL-QL, Classic' QL, …

RIF WG
AIR accountability tool

Govt metadata search
Linked Open Govt Data

SPARQL to Xquery translator

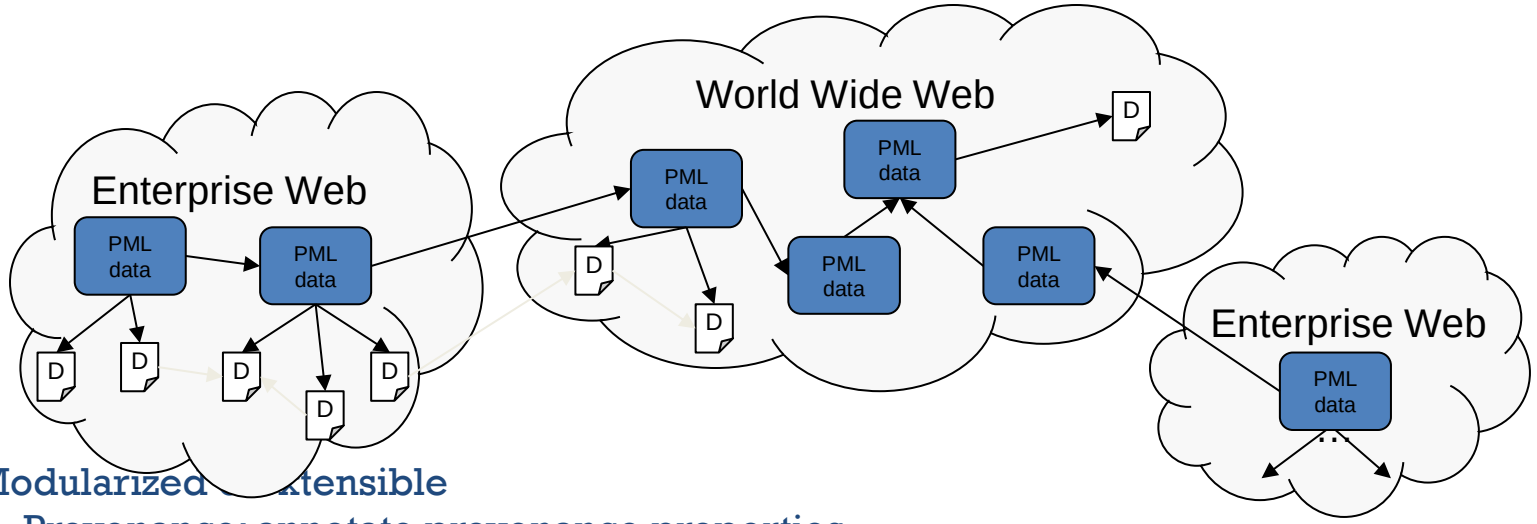
RDFS materialization (Billion triple winner)

Transparent Accountable Datamining Initiative (TAMI)

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Foundations: Proof Markup Language -> PROV

- A new kind of linked data on the Web

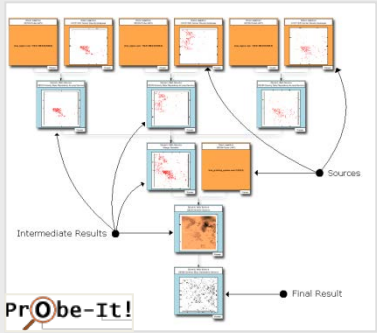


- Modularized & extensible
  - Provenance: annotate provenance properties
  - Justification: encodes provenance relations
  - Trust: add trust annotation
- Semantic Web based

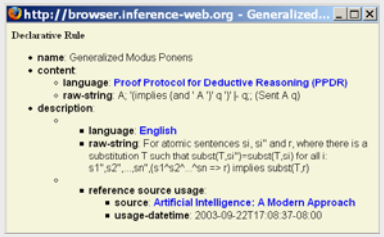
Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Foundations: Inference Web (IW)

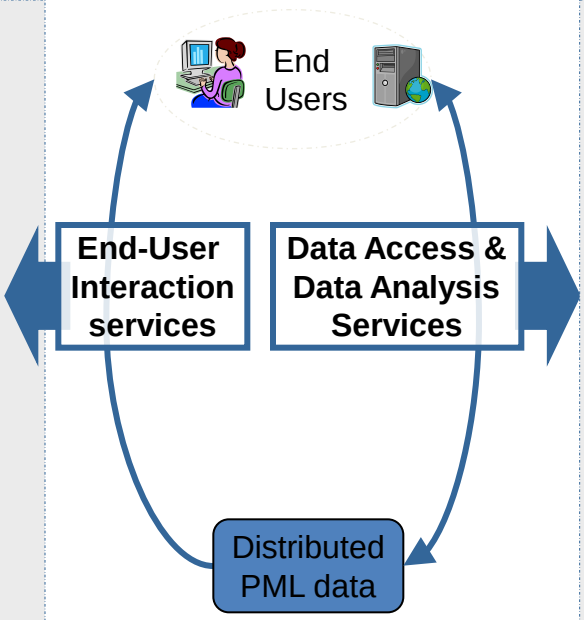
Explanation via Summary

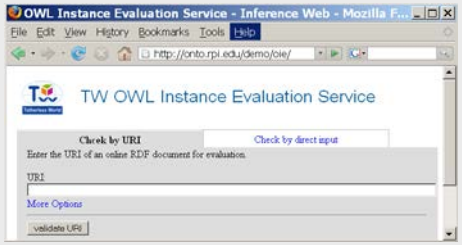


Explanation via Graph



Explanation via Annotation





Validate published PML data

Access published PML data

Inference Web is a semantic web-based knowledge provenance management infrastructure:

- PML for encoding and interchange provenance metadata in distributed environment
- Interactive explanation services for end-users
- Data access and analysis services for enriching the value of knowledge provenance

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Social Observatory

Social Media use is on the rise. Every day, we write:

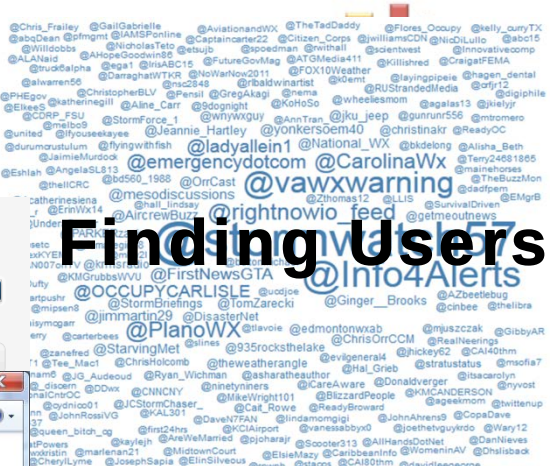
294 billion emails

2 million blog post

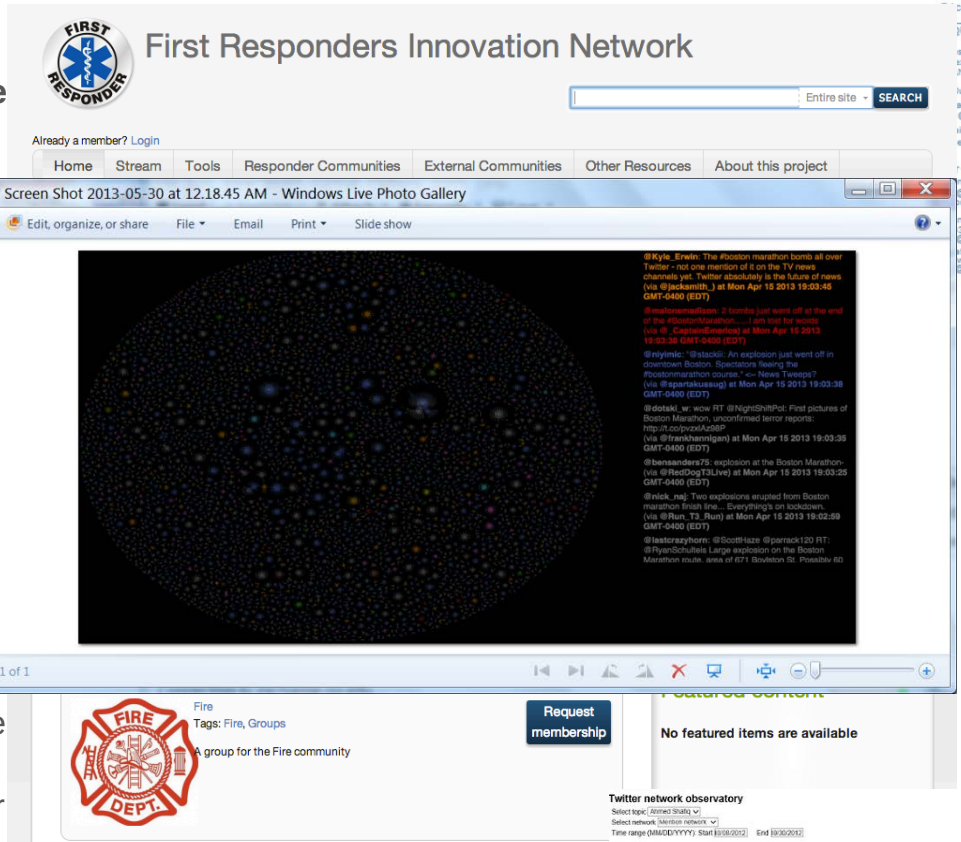
Over 40 Million Tweets*

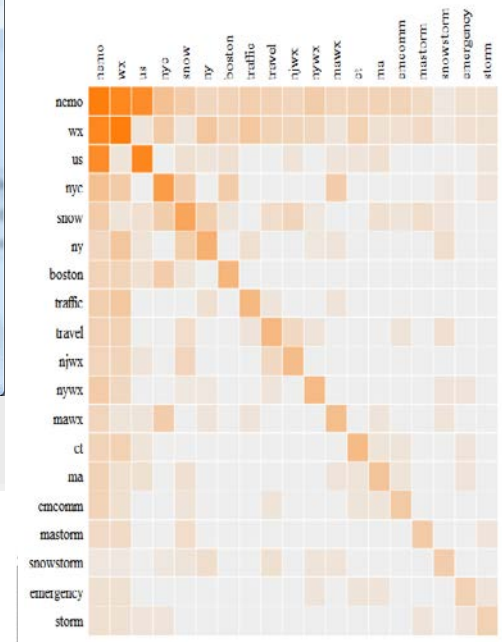
How can we leverage Social Media sites

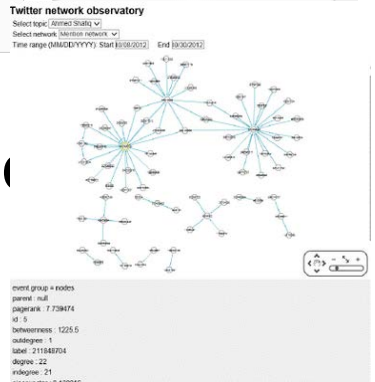
- to gather requireme for active First Responders?
- to identify communitie
- to identify stakeholder within those First Responder communities?
- To identify trends



Finding Users

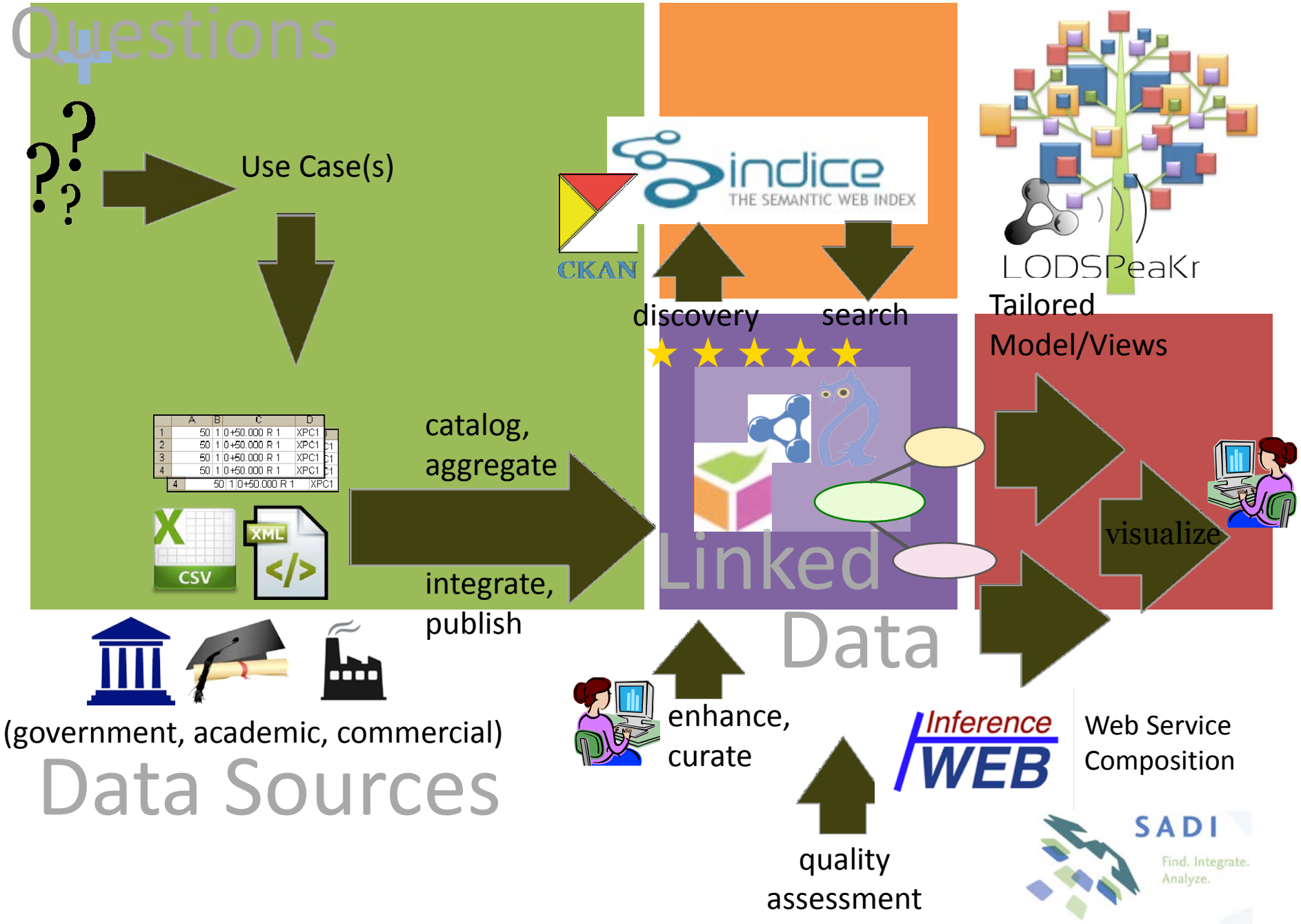






Finding Topi

http://tw.rpi.edu/web/project/FirstRes



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Reflections

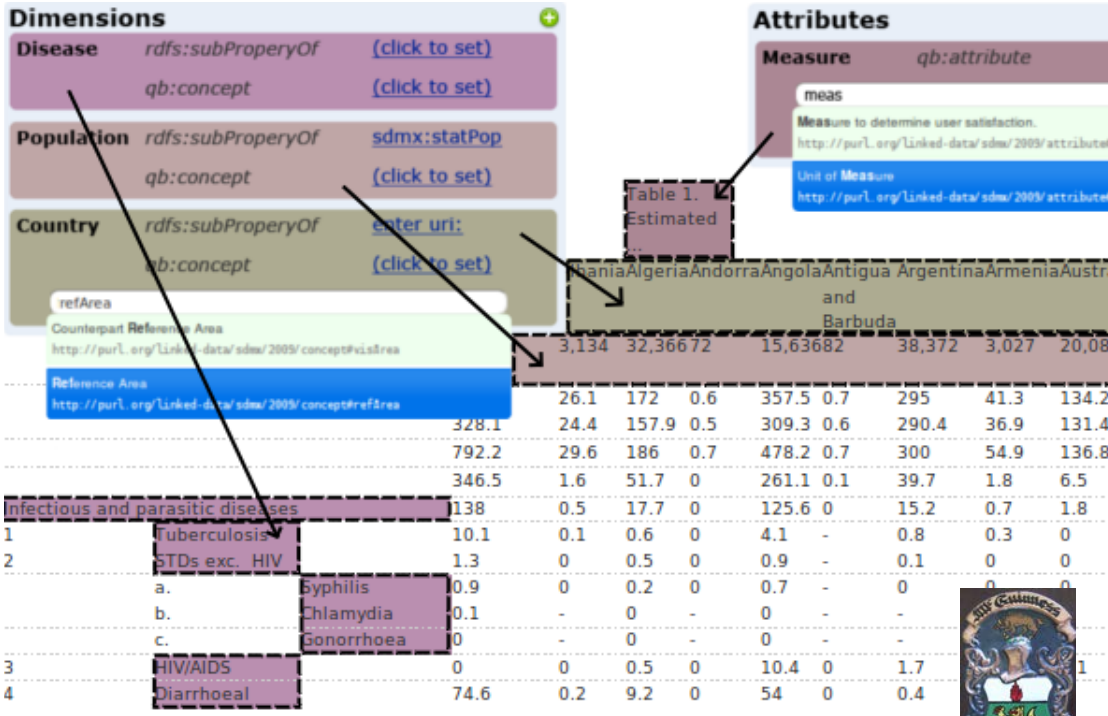
Successful but….

- What if we could allow data experts to build their own demos?
- What if we could allow non-subject matter experts to function as subject-literate staff?
- What if team members could interchange roles (and thus make contributions in other areas)?
- What technological infrastructure is required?
- Claim: all of this is being done now – but not at scale

Copyright © 2015 Thematix Partners LLC & McGuinness Associates

RDF Data Cube Vocabulary

- For publishing multi-dimensional data, such as statistics, on the web in such a way that it can be linked to related data sets and concepts using RDF.
- Compatible with the cube model that underlies SDMX (Statistical Data and Metadata eXchange).
- Also compatible with:
  - SKOS, SCOVO, VoiD, FOAF, Dublin Core Terms
- Integrated with the LOGD data conversion infrastructure
- Integrated with other tooling like Stats2RDF



Copyright © 2015 Thematix Partners LLC & McGuinness Associates

Community Health Status Indicators: DataCube Browser

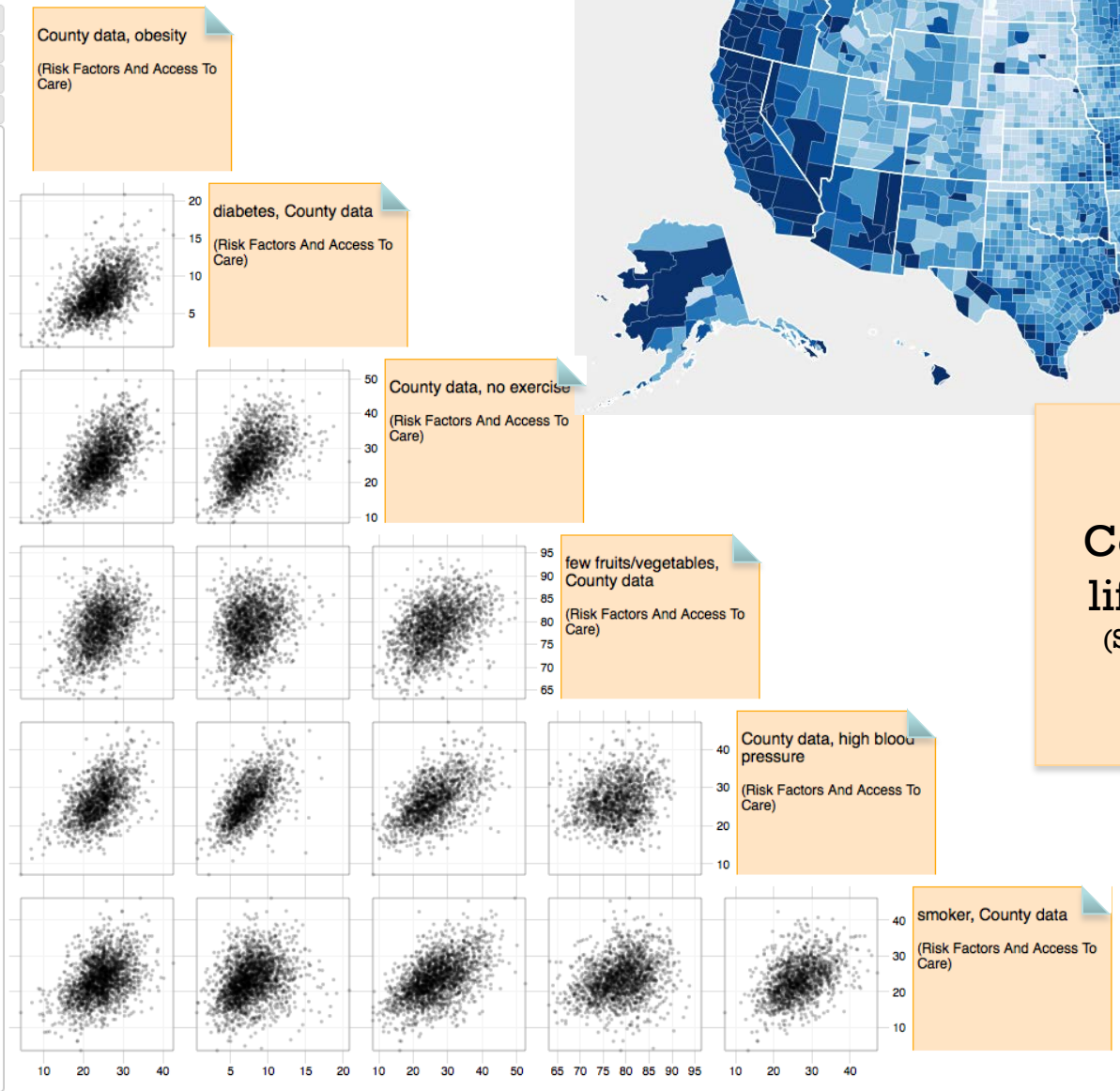
Measures

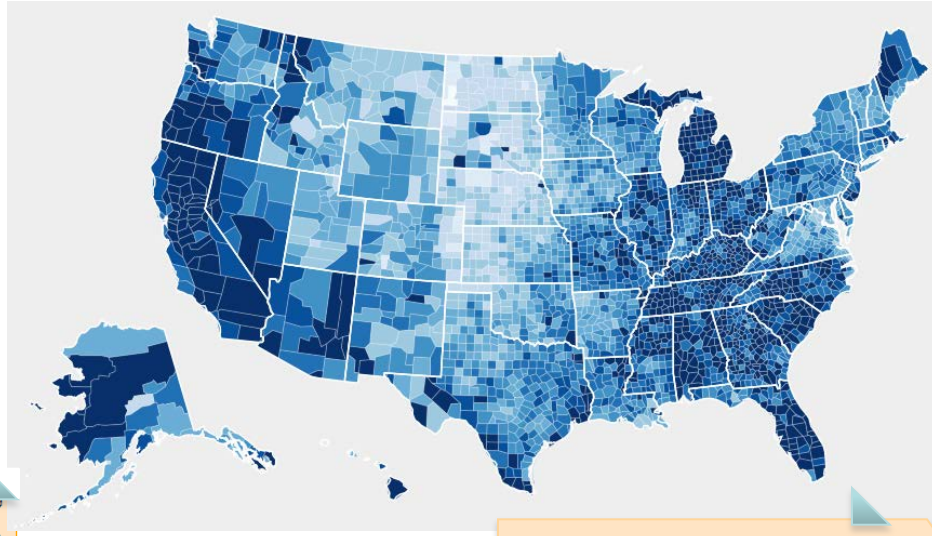
Select two or more measures.

- Relative Health Importance
- Summary Measures Of Health
- Demographics
- Preventive Services Use
- Risk Factors And Access To Care

Scatter	Map	Measure
☐	○	County data, uninsured individuals
☐	○	Indicator for single county designated Health Professional Shortage Area
☑	○	County data, obesity
☐	○	Indicator for any Community/Migrant Health Centers located in the county
☑	○	diabetes, County data
☐	○	000 pop., County data, dentists per 100
☑	○	County data, no exercise
☐	○	elderly (age 65+), medicare beneficiaries, County data
☑	○	few fruits/vegetables, County data
☐	○	medicare beneficiaries, County data, disabled
☐	○	000 pop., County data, primary care physicians per 100
☑	○	County data, high blood pressure
☑	○	smoker, County data

- Measures Of Birth And Death
- Vunerable Pops And Env Health
- Leading Causes Of Death





County average life expectancy
(Summary Measures of Health)

7/20/05