Semantics I

Matthew J. Graham
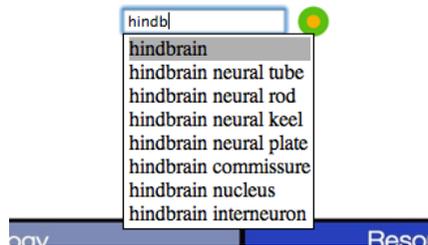CD^3

Methods of Computational Science
Caltech, 2015 April 9

matthew graham

- The future of the Internet (Web 3.0)

- Decentralized platform for distributed knowledge

- A web of databases

- Logical pieces of meaning that can be mechanically manipulated by a machine

- Vocabularies for making assertions about things

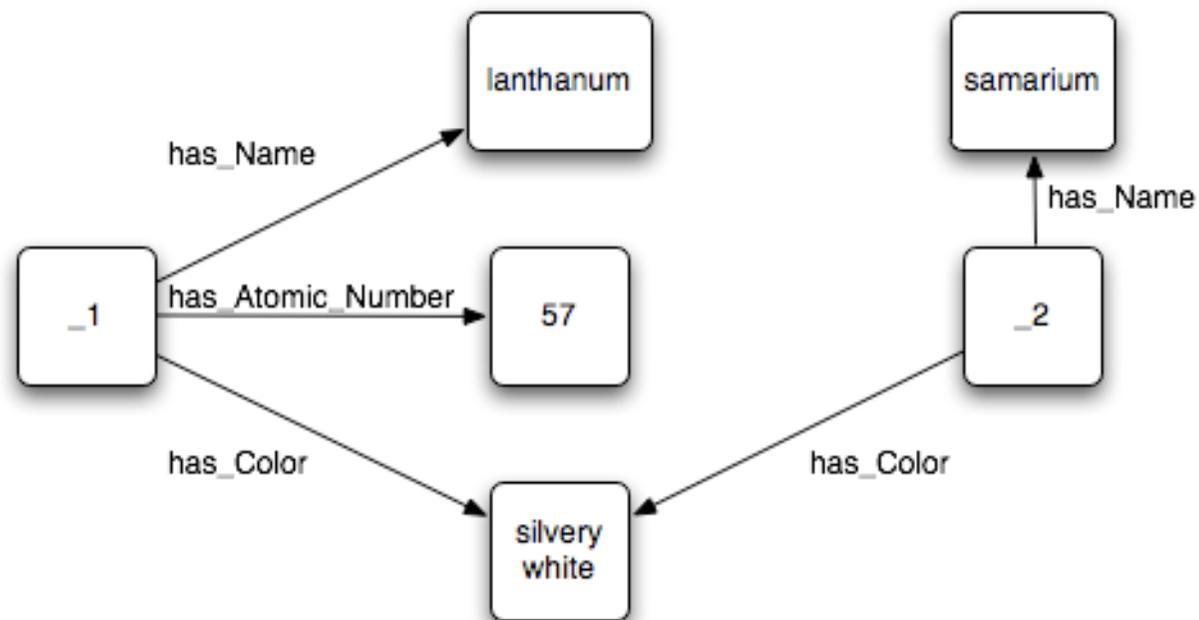- Data consistency and inferencing

- Smart applications

matthew graham

smart apps

Browse by anatomy/region

hindb

- hindbrain
- hindbrain neural tube
- hindbrain neural rod
- hindbrain neural keel
- hindbrain neural plate
- hindbrain commissure
- hindbrain nucleus
- hindbrain interneuron

Search criteria
anatomy: hindbrain and related terms

48 ensembl genes

name	description	ensembl id	alleles	matched term	score	
rbms3	RNA binding motif, single stranded interacting protein	ENSDARG00000044574	ct30a	hindbrain	1.0	view
			ct32a	hindbrain	1.0	view
rap1b	RAS related protein 1b	ENSDARG00000008867	ct40a	hindbrain	1.0	view
			ct48b	hindbrain	1.0	view
prkca	protein kinase C, alpha	ENSDARG00000039241	ct54a	hindbrain	1.0	view
abi1b	abl-interactor 1b	ENSDARG00000062991	ct67a, ct79a	hindbrain	1.0	view
			ct74a	hindbrain	1.0	view
sertad2	SERTA domain containing 2	ENSDARG00000055530	ct75a	hindbrain	1.0	view
syn2b	synapsin IIb	ENSDARG00000078971	ct122c	hindbrain	1.0	view
			ct168a	hindbrain	1.0	view
prkca	protein kinase C, alpha	ENSDARG00000039241	ct7a	hindbrain commissure	0.875	view
			ct48b	hindbrain commissure	0.875	view
syn2b	synapsin IIb	ENSDARG00000078971	ct122c	hindbrain commissure	0.875	view
aldoca	aldolase C, fructose-bisphosphate, a	ENSDARG00000057661	ct126a	cerebellum	0.875	view
			ct132a	cerebellum	0.875	view
zgc:153607	zgc:153607	ENSDARG00000026804	ct2b	whole organism	0.75	view
ctnna	catenin (cadherin-associated protein), alpha	ENSDARG00000036118	ct3a	whole organism	0.75	view

matthew graham

Information is best expressed as a labelled, directed graph (entity-attribute-value data model)

resource description framework

| W3C standard for encoding knowledge

| A fact is expressed as a *subject-predicate-object* triple or statement:

_1 has_Atomic_Number 57

| Subjects, predicates and objects are given as names for entities

| Names are URIs

| Objects can also be given as text values, called literal values, which may or may not be typed using XML Schema datatypes

matthew graham

n3/turtle:

```
@prefix pt: <http://www.example.org/PeriodicTable#> .
<http://www.example.org/PeriodicTable#La> pt:name "lanthanum" ;
    pt:atomicNumber 57 ; pt:color "silvery white" .
```

RDF/XML:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:pt="http://www.example.org/PeriodicTable#">
  <rdf:Description rdf:about="http://www.example.org/PeriodicTable#La">
      <pt:name> lanthanum <pt:name>
      <pt:atomicNumber> 57 </pt:atomicNumber>
      <pt:color> silvery white </pt:color>
    </rdf:Description></rdf:RDF>
```

matthew graham

```
<html xmlns="http://www.w3.org/1999/xhtml"
      xmlns:pt="http://www.example.org/PeriodicTable#">
  <head>
    <title> Lanthanum </title>
    <base href="http://www.example.org/PeriodicTable#La" />
  </head>
  <body>
    <p>
      The element <span property="pt:name"> lanthanum </span> has an atomic
      weight of <span property="pt:atomicWeight"> 57 </span> and is
      <span property="pt:color"> silvery white </span> in color.
    </p>
  </body>
</html>
```
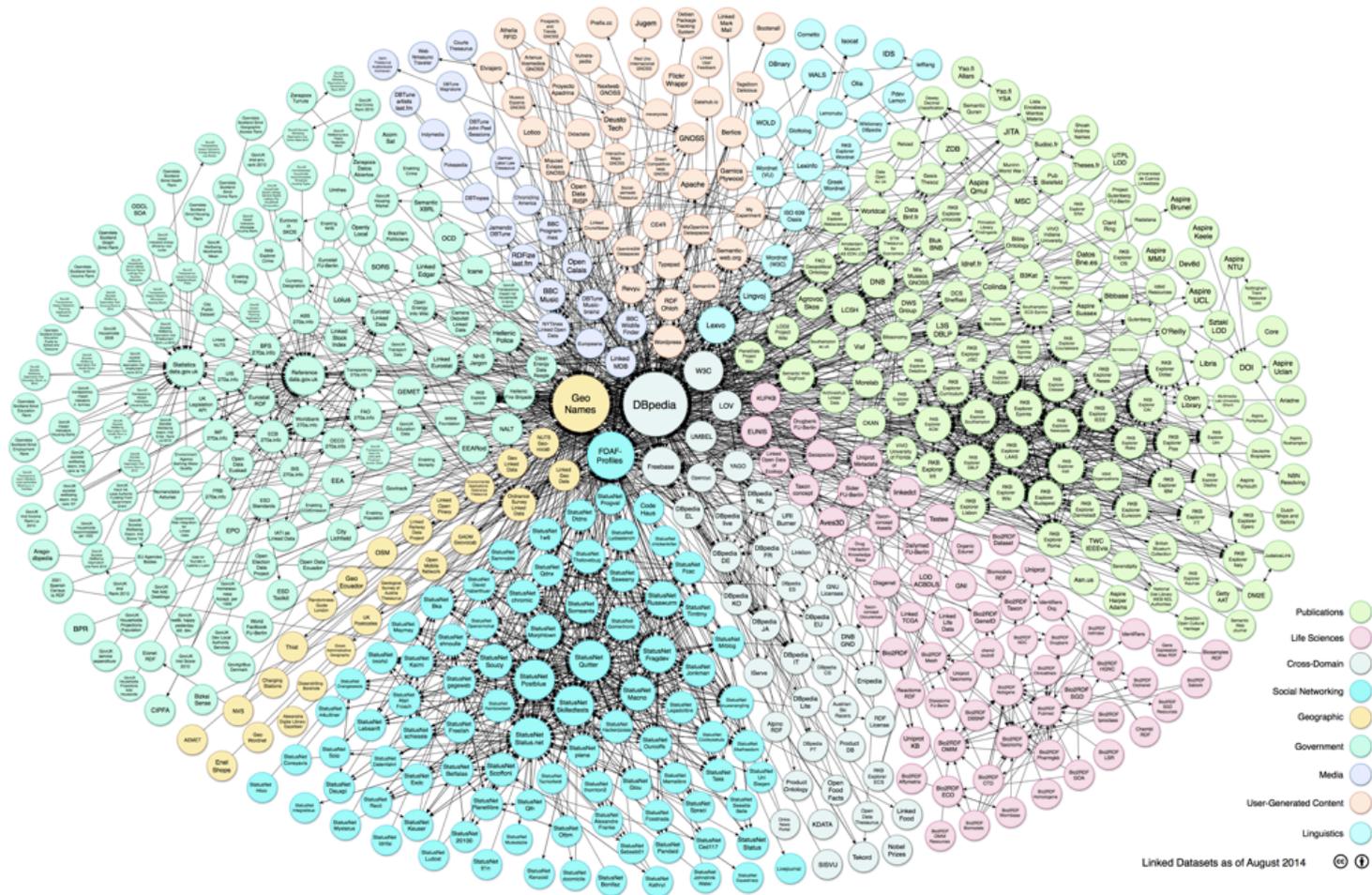
matthew graham

▌A term coined by Tim Berners-Lee who outlined four principles:

-- Use URIs to identify things that you expose to the Web as resources

-- Use HTTP URIs so that people can locate and look up (dereference) these things

-- Provide useful information about the resource when its URI is dereferenced

-- Include links to other, related URIs in the exposed data

Linked Datasets as of August 2014

Publications
Life Sciences
Cross-Domain
Social Networking
Geographic
Government
Media
User-Generated Content
Linguistics

matthew graham

▌List all episodes of the AMC television series "Breaking Bad" ordered by their air-date

▌Find the official websites of companies with more than 50000 employees

▌Find me things close to the Eiffel Tower

▌Discover new drugs to treat Alzheimer's:

-- What proteins are involved in signal transduction **and** are related to pyramidal neurons?
    * Google: 223000 hits, 0 results
    * Linked healthcare data: 32 hits, 32 results

matthew graham

Browsers:

-- Tabulator (also has Firefox plugin)
-- Disco
-- Openlink Data Browser

Libraries

-- Semantic Web Client Library (Java)

Exposing data

-- d2r for exposing relational data in a db

matthew graham