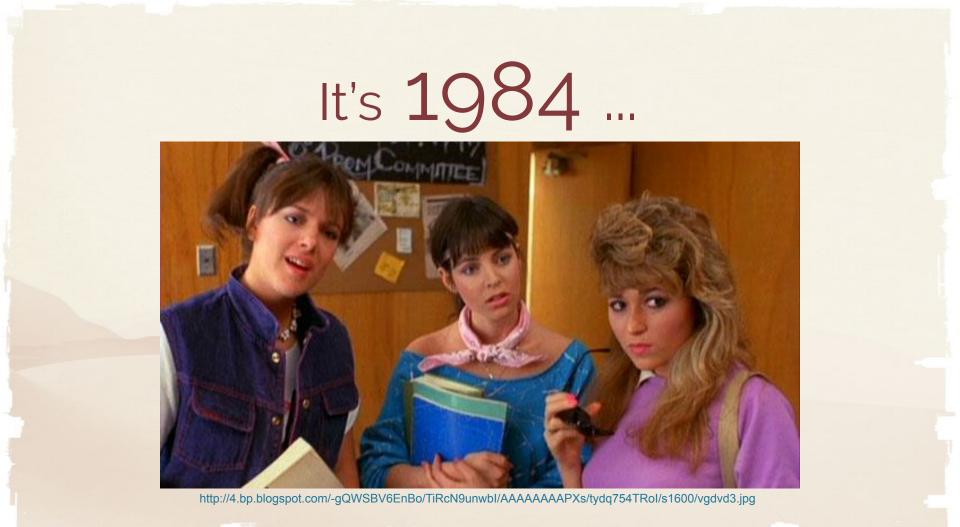
Use of Controlled Vocabularies

Potential applications to time series data

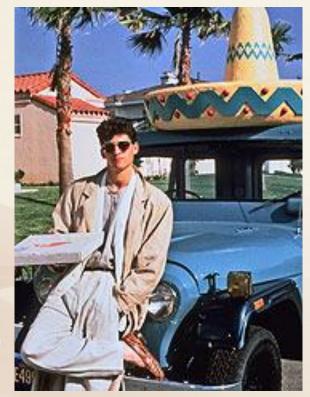


Friday night



http://i.telegraph.co.uk/multimedia/archive/01594/corey-haim_1594267c.jpg

MISSION: PIZZA



"What's their address?"

"1428 Elm St."



"Be there in 30 mins or less

...or it's free."

http://img2.timeinc.net/people/i/2006/celebdatabase/patrickdempsey/patrick_dempsey5_180_240.jpg http://img.timeinc.net/time/daily/2009/0901/360_elm_street_0105.jpg

MISSION: MOVIE

"GET THE ONE WITH..."

Righteous girl + bodacious dude

Light beams flying everywhere

The one with the "Force"



http://swhomevid.site40.net/vhs_anh1.jpg

The search begins...





https://deadlymovies.files.wordpress.com/2012/01/photo-2.jpg

MOVIE: FOUND!

Righteous girl + bodacious dude

✓ light beams flying everywhere

✓ The one with the "Force"



http://www.chud.com/wp-content/uploads/2012/07/MEGAFORCE-FRONT-copy.jpg

BACK AT THE HOUSE...



http://static.guim.co.uk/sys-images/Film/Pix/pictures/2010/2/24/1267010697702/Winona-Ryder-Kim-Walker-L-001.jpg

A Tale of Two Elm Streets





Which 'Elm St' was that?



ELM ST. near Austin, TX



http://brooklynsteez.com/products/square/81690.png http://tinyurl.com/mtmlbpx

At least, you have MEGAFORCE

✓ Righteous girl + bodacious dude

✓ light beams flying everywhere

✓ The one with the "Force"



http://www.chud.com/wp-content/uploads/2012/07/MEGAFORCE-FRONT-copy.jpg

Betamax vs. VHS





http://images.sodahead.com/slideshows/000000228/betamax-11979413784_xlarge.jpeg http://upload.wikimedia.org/wikipedia/commons/thumb/3/3c/Betavhs2.jpg/220px-Betavhs2.jpg

WHAT HAPPENED, DUDE?

PIZZA no results = *ambiguous* address

MEGA FORCE got results, not what was *expected*

BETAMAX results weren't *useful*

Findability and Accessibility has greatly improved.



Scaling interoperation is difficult

F.A.I.R.



Without progress, ocean data risks being siloed

What does interoperation require?

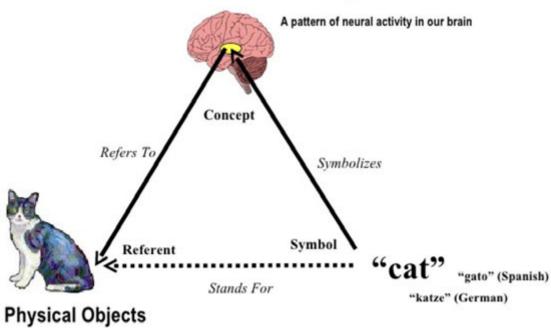
AGREEMENT





Of Concepts & Computers

Semantic Triangle



Referents MUST pass through Concepts

Computers can only exchange *symbols*

People use **context** to derive the correct meaning

Ogden, C. K., & Richards, I. A. (1923) The Meaning of Meaning

McReary (2006) Patterns of Semantic Integration CC BY 2.5

SEARCH: "COASTAL BEEF"



Leftcoast 25lb assortment

An assortment of quality, grass-fed beef cuts steaks, roasts, and stir fry/fajita cuts... Read more >

>

Beef Recall Update: Where we stand. 02/18/14

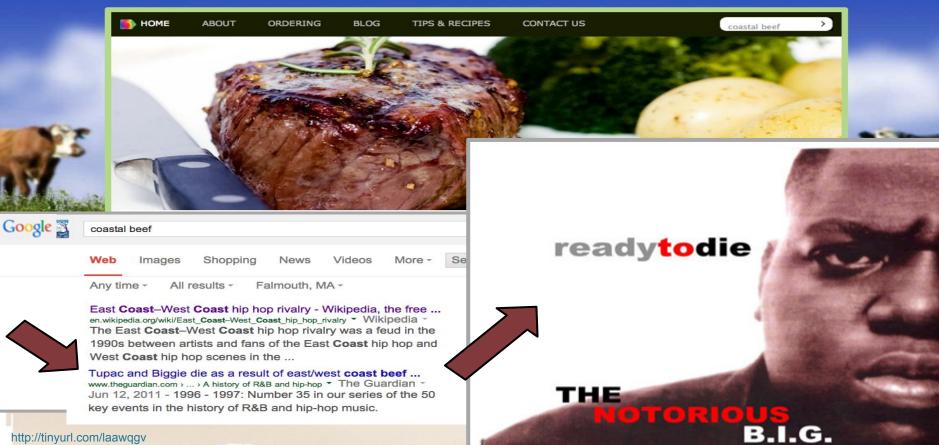
Dear friends, customers and consumers,

<

Like all our fellow sustainable ranchers affected by the Rancho recall, we have been trying to glean all information and build a strong, coherent response. At this point, we don't have any information from USDA about what happened. LeftCoast GrassFed and our hardworking, dedicated colleagues-small and family owned ranches, the supporters of rural America-have been struggling to ascertain if our beef was affected by the allegations made by the USDA against Rancho. The USDA's original recall was focused on one day (January 8, 2014) and, as we understand it, only on culled dairy cows that were processed without full inspection.

http://www.leftcoastgrassfed.com/

IS "COASTAL BEEF" DEADLY?



http://i343.photobucket.com/albums/o465/MattLeader/Biggie-ReadyToDie-CD.jpg

<name>Adam Shepherd</name> <address>1428 Elm St.</address> <city>Austin</city> <state>TX</state>

<名称>亚当谢泼德</名称> <地址>1428 榆樹街</地址> <市>奧斯汀</市> <國>德州</國>

Without a "data dictionary" it's difficult to know that meaning of the data elements is.

Tags appear in patterns, but meaning is a mystery to a computer

<name>

<名称>

https://schema.org/name

Machines can detect if two URIs are the same.

McReary (2006) Patterns of Semantic Integration CC BY 2.5

https://schema.org/name

Machines can detect if two URIs are the same.

https://orcid.org/0000-0003-4486-9448

亚当谢泼德

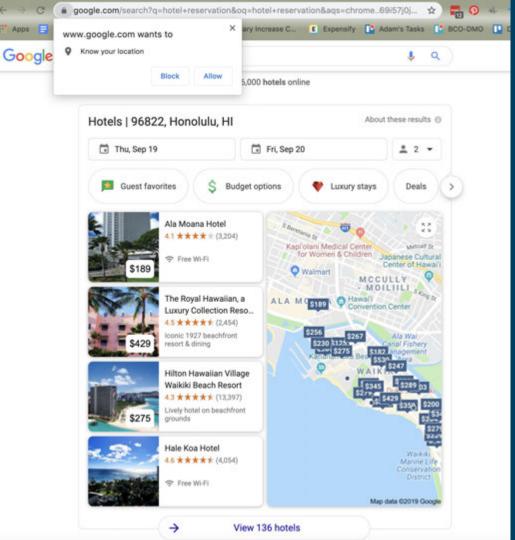
Adam Shepherd

<name>

<名称>

McReary (2006) Patterns of Semantic Integration CC BY 2.5

Works now, today



Data on the Web Best Practices

W3C Recommendation 31 January 2017



DISCOVERABILITY

Provide metadata

Provide descriptive metadata

Use persistent URIs as identifiers of datasets

Use persistent URIs as identifiers within datasets

Assign URIs to dataset versions and series

Use Web Standards as the foundation of APIs

Cite the Original Publication

COMPREHENSION

Provide metadata Provide descriptive metadata Provide structural metadata Provide data provenance information Use locale-neutral data representations Reuse vocabularies, preferably standardized ones Choose the right formalization level Gather feedback from data consumers

Enrich data by generating new data **Provide Complementary Presentations**

ACCESS Provide bulk download Provide Subsets for Large Datasets Use content negotiation for serving data available in multiple formats Provide real-time access Provide data up to date Make data available through an API Use Web Standards as the foundation of APIs

Provide Complementary Presentations

INTEROPERABILITY

Use persistent URIs as identifiers of datasets

Use persistent URIs as identifiers within datasets

Reuse vocabularies, preferably standardized ones

Choose the right formalization level

Make data available through an API

Use Web Standards as the foundation of APIs

Avoid Breaking Changes to Your API

Provide Feedback to the Original Publisher

W3C Data on the Web Best Practices

INTEROPERABILITY

Use persistent URIs as identifiers of datasets

Use persistent URIs as identifiers within datasets

Reuse vocabularies, preferably standardized ones

Choose the right formalization level

Make data available through an API

Use Web Standards as the foundation of APIs

Avoid Breaking Changes to Your API

Provide Feedback to the Original Publisher

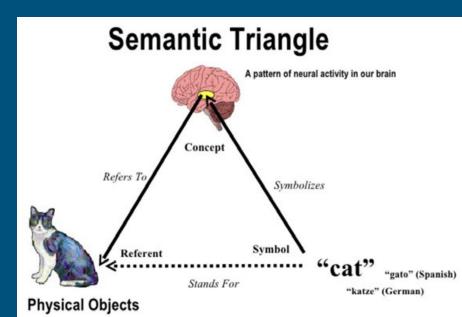
https://www.w3.org/TR/dwbp/#challenges

Making ocean data agreeable (with each other)?

AGREEMENT on identifiers within

datasets

- Parameters
- Methods
- Instrumentation
- Etc...









DATABASE

Programs	44	Para
Projects	1,085	Short I Short N
Deployments	2,888	Official
Platforms	596	Units: Units E
Datasets	9,447	Grapha No Dat
Instruments	487	Maxim
Parameters	1,420	» skos:
People	2,736	Minimu
Affiliations	594	- Dese
Funding	93	The
Awards	2,044	

GEOSPATIAL ACCESS



Parameter: Magnetic_susceptibility

Short Description: Magnetic_susceptibility
Short Name: Magnetic_susceptibility
Official Name: Magnetic susceptibility
Units:
Units External Identifier:
Graphable:
No Data Value:
Maximum Value:
External Identifier:
» skos:exactMatch http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/ C*
Minimum Value:

Description

The degree to which a sediment sample is affected by a magnetic field.

Datasets associated with this parameter

Dataset Name 🝝	Brief Description	Project	PI-Supplied Parameter Name
Palau Lake Core Properties	Core Logger Physical Properties for Palau Lakes Sediment Cores	PaPaPro	MSCL_MS_Loop

.



Q

DATABASE

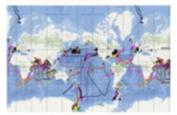
Programs	Programs	44
	Projects	1,085
	Deployments	2,888
	Platforms	596
	Datasets	9,447
	Instruments	487
Pe	Parameters	1,420
	People	2,736
	Affiliations	594
	Funding	93
	Awards	2,044

Parameter: Magnetic_susceptibility Short Description: Magnetic_susceptibility Short Name: Magnetic_susceptibility Official Name: Magnetic susceptibility Units: Units External Identifier: Graphable: No Data Value: Maximum Value: External Identifier: * skos:exactMatch http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/ Minimum Value:

Description

The degree to which a sediment sample is affected by a magnetic field.

GEOSPATIAL ACCESS



Datasets associated with this parameter

Dataset Name 📥	Brief Description	Project	PI-Supplied Parameter Name
Palau Lake Core Properties	Core Logger Physical Properties for Palau Lakes Sediment Cores	PaPaPro	MSCL_MS_Loop

.

http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/

1 -- Magnetic susceptibility of the sediment --

URI	http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/
Identifier ()	SDN:P01::MAGSSEDM
Preferred label (en)	Magnetic susceptibility of the sediment
Alternative label (en)	MagSus_Sed
Definition (en)	The degree to which a sediment sample is affected by a magnetic field.
Version Info ()	1
Has Current Version	http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/1/
PAV Version ()	1
PAV Authored On ()	2010-11-04 16:17:23.0
Deprecated()	false
Broader	http://vocab.nerc.ac.uk/collection/P02/current/XMGS/
Broader	http://vocab.nerc.ac.uk/collection/S26/current/MAT00136/
Related	http://vocab.nerc.ac.uk/collection/P06/current/UCGS/
Related	http://vocab.nerc.ac.uk/collection/S02/current/S032/
Date ()	2010-11-04 16:17:23.0

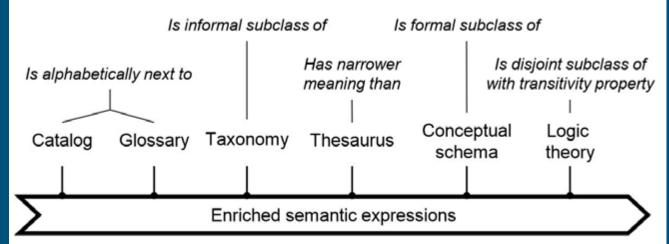
http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/

	a successful the section and	_
∽ №	xml version="1.0" encoding="UTF-8"? xml-stylesheet href="/VocabV2/Concept2Html.xsl" type="text/xsl" media="screen"? <rdf:rdf <="" th="" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:skos="http://www.w3.org/2004/02/skos/core#"><th></th></rdf:rdf>	
	<pre>xmlns:rdc="http://purl.org/dc/terms/" xmlns:dce="http://purl.org/dc/elements/1.1/"</pre>	
in the second	xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:owl="http://www.w3.org/2002/07/owl#"	
URI	xmlns:void="http://rdfs.org/ns/void#" xmlns:pav="http://purl.org/pav/" xmlns:prov="https://www.w3.org/ns/prov#"	
T 1	<pre>xmlns:reg="http://purl.org/linked-data/registry#" xml:base="http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/"> <skos:concept rdf:about="http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/"></skos:concept></pre>	
Identifi	<pre><skos:coreflabel xml:lang="en">Magnetic susceptibility of the sediment</skos:coreflabel></pre>	
Duchan	cekeesal#Tabel wml:langs"on"-MagCus Code/ekeesal#Tabel>	
Preferre	<pre><skostarthaber xmtriang="en">Ragaus_sed</skostarthaber>The degree to which a sediment sample is affected by a magnetic field. </pre>	
and the second sec	<dc:date>2010-11-04 16:1/:23.0</dc:date>	
Alterna	<dc:identifier>SDN:P01::MAGSSEDM</dc:identifier>	
Definiti	<skos:notation>SDN:P01::MAGSSEDM</skos:notation>	tic field.
Deminu	<owl:versioninfo>1</owl:versioninfo>	tic neia.
Version	<pre><pav:hascurrentversion rdf:resource="http://vocab.nerc.ac.uk/collection/P01/current/MAGSSEDM/1/"></pav:hascurrentversion></pre>	
VEISION	<pre><pav:version>1</pav:version> <pav:authoredon>2010-11-04 16:17:23.0</pav:authoredon></pre>	
Has Cu		
Has Cu	<owl:deprecated>false</owl:deprecated>	
PAV Ve	<skos:broader rdf:id="I457489" rdf:resource="http://vocab.nerc.ac.uk/collection/P02/current/XMGS/"></skos:broader>	
1710 00	<skos:related rdf:id="I457491" rdf:resource="http://vocab.nerc.ac.uk/collection/P06/current/UCGS/"></skos:related> <skos:broader rdf:id="I607358" rdf:resource="http://vocab.nerc.ac.uk/collection/S26/current/MAT00136/"></skos:broader>	
PAV Au	<pre><skos:related rdf:id="I632964" rdf:resource="http://vocab.nerc.ac.uk/collection/S02/current/S032/"></skos:related></pre>	
	<void:indataset rdf:resource="http://vocab.nerc.ac.uk/.well-known/void"></void:indataset>	
Depreca		
and the second se	<rdf:description rdf:about="#I457489"> <prov:has provenance="" rdf:resource="http://vocab.nerc.ac.uk/mapping/I/457489/"></prov:has></rdf:description>	
Broade		
	crdf. Description rdf. about ##1457491">	
Broade		
Delated		
Related	<rdf:description rdf:about="#I607358"> <prov:has_provenance rdf:resource="http://vocab.nerc.ac.uk/mapping/I/607358/"></prov:has_provenance></rdf:description>	
Related		
Related		
Date ()	<prov:has_provenance rdf:resource="http://vocab.nerc.ac.uk/mapping/I/632964/"></prov:has_provenance> 	
Date ()	 	

Controlled Vocabularies

• A list of terms and definitions <u>identified by URIs</u>

- Some governance model for ensuring consistency
- Defined in relation to each other



https://tw.rpi.edu/weblog/tag/ontology-spectrum/

Making ocean data agreeable (with each other)?

Recommendation	Solution	Where is the Solution?
Identify datasets with a persistent URI.	DOIs	Data Repository*
Identify dataset <u>contents</u> with a persistent URI.	Controlled Vocabulary URIs	Domain Repository
Reuse vocabularies, preferably standardized ones.	Community	You!

https://www.the-scientist.com/news-opinion/the-push-to-replace-journal-supplements-with-repositories--66296

In Review

- 1. We need AGREEMENT about dataset contents.
- 2. Agreements materialized as symbols (URIs) for computers.

Questions?

EXTRA SLIDES

Choosing a Controlled Vocabulary

- Availability
- Quality
 - Governance Model
 - Clarity and Precision
 - Completeness
 - Encoding Format (HTML v. Excel/CSV v. RDF/OWL)
- Community Adoption

Works now, today...even for Datasets

Google Dataset Search

nitrogen "Gulf of Maine"

× About

23 results found

Carbon and nitrogen content of E. huxleyi at 3 pCO2 levels, 2011... www.bco-dmo.org

Nutrient data from the western

Phytoplankton chlorophyll and

Model output from nutrients-

phytoplankton-zooplankton-...

archaeological fish bones from

nutrient studies from R/V...

Gulf of Maine (bottles) collecte...

Published Dec 13, 2016

www.bco-dmo.org

www.bco-dmo.org

www.bco-dmo.org

Published Dec 18, 2013

C and N isotope data of

Published Jun 1, 2015

Published Jul 23, 2010

Carbon and nitrogen content of E. huxleyi at 3 pCO2 levels, 2011-2012 (E Hux Response to pC project)

Emiliania huxleyi CN content

Explore at www.bco-dmo.org

Dataset published Dec 13, 2016

Dataset provided by **Biological and Chemical Data Management Office**

Authors

Tristen Wuori; Suzanne Strom; Dr Brooke Love; Dr Brady M. Olson

License

https://creativecommons.org/licenses/by/4.0/

Available download formats from providers

csv, vnd.datapackage+json

Variables measured

CO2 treatment, Days of semi-continuous culture, E. huxleyi strain numbe Nitrogen per cell in picograms, Particulate inorganic carbon per cell in picograms, Particulate organic carbon per cell in picograms, Sample date replicate, Total carbon per cell in picograms

Dataset funded by

NSF Division of Ocean Sciences