Semtools

for the semantic annotation and querying of environmental data

Ben Leinfelder, Mark Schildhauer, Shawn Bowers, Huiping Cao, Matt Jones, Margaret O'Brien

SONet/Semtools/Data Conservancy/DataONE linkup meeting Ithaca, July 20-21, 2010

Monday, July 19, 2010

Data challenges in the environmental sciences

- multidisciplinary data: ecology, genomics, climatology, hydrology, soil, etc.
- many data formats, many different types of measurements: huge heterogeneity
- synthetic, integrative science

- How to discover and interpret data needed for integrative, synthetic environmental science?

- metadata and keywords are not enough: ambiguous, idiosyncratic
- controlled vocabularies: a good start, but can do more with today's technology
- *ontologies: based on Web standards, provide power of inferencing

Observations & Measurements

Many environmental data sets are stored in tables (relations):

columns contain some consistent measurement type, rows represent multiple recorded values of those measurements, and are associated in a tuple due to some "contextual" (Plot contains Plant) or other relationship (Biomass and Height of same Plant)

Must accommodate table structures!

...but also vectors, rasters, matrices, more...

How to use the SONet core model?

- Data annotation tool
- OM query language
- OM framework for data discovery

loc	quad	nitr	wt	place	treat	plot	LL
SCAL	1	N	6.2	Sth	C	1	0.003
SCAL	2	Y	7.2	Sth	C	1	0.002
CCAL	1	N	4.2	Sth	N	1	0.008



Monday, July 19, 2010







Morpho



- documents ecological data through formal metadata
- based on Ecological Metadata Language (EML)- XML-schema
- local and network storage and querying
- supports attribute-level descriptions of tabular data
- originally developed under NSF-funded KNB project

Semtools



Extends Morpho codebase

- builds on existing rich metadata corpus (KNB)
- semantic annotation of data through metadata
- map attribute-level metadata descriptions to observation model
- uses core model defined by SONet
- access domain ontologies through OBOE
- semantic querying

Open Data Annotation Frame

ne Player File Edit View Share Window Help	● Stop Recording 49 📕 💲 🛜 40 (■ (2:4	7) Tue 8:17 AM
File Edit Search Decumentation Data Appetation Window Hele		
File Edit Search Documentation Data Annotation Window Help	30	
	**	
Current profile: benriver		Macintos
(uid=leinfeider,o=NCEAS,dc=ecoinformatics,dc=org)		
Change profile: benriver :		
	were first the second	
Create a new prome Weicome to Morpho	0!	
Network Status: NOT Looped Te		
If you do not choose to login, you will be able to access only		
"public" network files as a Guest User		
Login to network using current profile:		everyth
Password: (login)		
*- Work with your data:		
E, Create a new data package		
Cpen an existing data package		
Search for an existing data package		
	- O -	
	and the second se	

🛋 Q

Semantic Annotation

Apply semantic annotation to data attribute of "veg_plant_height"

- Characteristic (Height)
- Entity (Plant)
- Standard (Meters)

term from Observation Ontology (OBOE.OWL) term from Domain Ontology (Plant-trait.OWL)

Semantic Annotation

ne Player	File Edit Vie	w Share Wir	dow Help	-				-	Stop Recording 🕤 📕 💈 🛜	4)) 💽 (1:31) 💽	Tue 8:31 AM 👤
	0.00				Data F	ackane: henris	ar 202.4				
	Eile Edit (learch Decum	entation Dr	Annoti	tion Wind	ow Help	61-606-4			_	
	File Edit 3	search Docum	ientation Da	ita Annota	tion wind	ow Help				- 37	
		084 6	8								
	Marin Angelian	alanasta: Trait	data							Taxan International Control of Co	
	Accession Nun	nber: benriver.20	2.4 Keywords: p	lant, trait						3 local	Macintosh
	more										
	* 					<u></u>					
	-	Column Annotati	on Context	Annotation	Full Annota	tion Graph	Annotation		Metadata		
	Selected Attr	ribute: <none select<="" td=""><td>ted></td><td></td><td>s Distinct?</td><td>Is Key? Hel</td><td>p (</td><td>Edit</td><td>Entity/Attribute</td><td>hide X</td><td></td></none>	ted>		s Distinct?	Is Key? Hel	p (Edit	Entity/Attribute	hide X	
									Selected column or entity me	tadata	
			U	se pre-configur	ed measureme	nt template:					
		The Characteristic	of the Intity wa	at recorded	using the Sta	ndard	and the Proto	col	Entity Description		
									Name: OldFields ForMarie.cs	100 B	everythin
									Online Distribution Info:		
	Maximum	ate unce taken on th	a come comple o	e leadhaideast as -				(A)	Download File: ecogrid://knb/benriver.203.1		
	Measureme	nts were taken on tr	se same sample or	individual as:					Physical Structure Description:		
								- 1	Object OldFields_ForMarie.csv		
			OI	dFields_ForM	arie.csv				Size: 4063 byte		
		Lext	CEAL	molarity	number	dimensionless	molarity	molesi	Number of		
		SITE_ID ER_HCM	ACRIEUPA	38.08333	Ldeltal3C	_1.875	0.334	13390	Lines:		
		FR-HGM	ARENSERP	32.91666	-28.94	0.94	0.129	1.328	Text Record #x0A		
		FR-HGM	ARISROTU	0			0	0	Format: Attribute		
		FR-HGM	AVENBARB	34.45833	-28.69	0.237	0.265	1.360	Orientation: column		
	Data	FR-HGM	BRACPHOE	36.39583	-26.5675	-1.19	0.397	1.041	Simple Field		
		FR-HGM	BROMEREC	36.20833	-27.12	-3.75	0.3615	1.046	Number Of		
		FR-HGM	BROMMADR	35.33333	-28.255	0.28	0.2695	1.596	Records: 40		
		FR-HGM	CALANEPE	39.75	-29.995	0.065	0.235	1.582			
		2	CENTRALINE	se mere		0.03		1.12			
					_			1.1.1.1			
					Old	Fields_ForMari	e.csv				

₽

Query Precision

Keyword-based search

- "kelp"
- 3 data sets found

Observational semantics-based search

- Entity="kelp"
- I data set found

Query Precision, & Annotation Refinement

		ile Edit View Share Window Help	Stop Recording	• 🔲 🕴 🚖 🔹 (Charged) 🖬 Tue 5:33
Image: Image	Image: Bartanian Barta Annotation Window Help			
Ide Edit Search Documentation Data Annotation Window Help Ide Edit Search Documentation Control Ide Edit Search Documentation Data Annotation Window Help Ide Edit Search Documentation Control Ide Edit Search Documentation Data Annotation Window Help Ide Control Ide Control<	Morpho Morpho <th></th> <th></th> <th></th>			
		File Edit Search Documentation Data Ann	Morpho notation Window Help	
<complex-block></complex-block>				
<form></form>	Image: A company of the second sec			
<form></form>		Current profile: benriver	and the second s	Ma
Change profile: ● nentrer Create a new profile: Outcome for an existing data package Point an existing data package Open an existing data package Point an existing data package Point an existing data package		(uid=ieinfelder,o=NCEAS,dc=ecoinformatics,dc=org)	The second	
 Create a new profile Mort Actavity: NOT Logged In the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- the otor choose to logr, you will be able to access out- will be able to accession out- able to accession out- be able to accession out- able to accession out- be able to accession out- able to accession out- accession out- accession out- accession out-	ged In It be able to access only It copin	Change profile: benriver \$	the second state of the second state of the	
Itemark Status: NDT Logged In Type do not choose to login, you will be able to access only used. Login: Desword: Coate a new data package Coate a new data package Coate a new data package Search for an existing data package	ged 1n will be able to access only rent profile: Logn Blan kkagen: ba package	Create a new profile	Welcome to Morpho!	
Work with state a set use: Work with your data: Create a new data package Create a new sisting data package Search for an existing data package	Second Interpretation	Natural Clause III NOT Langed To	The second seco second second sec	
Public' network files as a Guet User Login to network using current profile: Pasword: Common according to the package Popen an existing data package Popen an existing data package Parcent for an existing data package		If you do not choose to login, you will be able to access only		
Login to network using current profile: Password: Create a new data package Open an existing data package Search for an existing data package Search for an existing data package		"public" network files as a Guest User	Line and the second and the second	
Password: Work with your data: Create a new data package Open an existing data package Search for an existing data package	Logn Rane Rapackage	Login to network using current profile:	The second	
Work with your data: Create a new data package Cpen an existing data package Search for an existing data package 	Re Ickage	Password: Logn	PLANS - THE PARTY OF THE PARTY	
 Create a new data package Open an existing data package Search for an existing data package 	Gew Ickage Ita package	Wark with your data-		and the second
 Contract first faith package Search for an existing data package 	ackage Ita package	P. Create a new data package	Land Marine Stream Print Billion	
Search for an existing data package	avage ta package	es Onen an existing data parkage	Transation of the second	
		Coper an existing data peckage	and the second s	
		er search for an existing data package	contraction and a first product of product of the	
				Alexandra II
ਦ O o ∠				
- C O =				
				₩ 0 m /

QuickTim

Query Expansion

Entity=Kelp AND Characteristic=DryMass

- I record
- Macrocystis is subclass of Kelp

Entity=Kelp AND Characteristic=Mass

- 2 Records
- DryMass is subclass of Mass

Query Expansion

QuickTime Player File	e Edit View Share Window Help	Stop Recording 🕤 💻	홍 🔿 🜒 💽 (Charged)	🖬 Wed 11:15 AM 👤 Q
	C Morpho File Edit Search Decumentation Data Association Window Hele			1. 1. 1. 1. 1.
1.1.1.1	File Edit Search Documentation Data Annotation Window Help			
			**	
	Current profile: benriver			
	(uid=leinfelder,o=NCEAS,dc=ecoinformatics,dc=org)			Mathinosi Ho
	Change profile: benriver \$			
	Create a new profile Welcome to Morphe	o!		Untitled
· · · · · · · · · · · · · · · · · · ·	Network Status: NOT Logged In			
	If you do not choose to login, you will be able to access only			
	"public" network riles as a Guest Oser			augusting .
	Partment:			everything
	Password.			
	Work with your data:			
	Create a new data package			
	Conen an existing data parkane			
	Control extra regions			
	er search for an existing data package			
				10 1 N 1
	and a second			
			- 0 r	
			A REAL PROPERTY OF	1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1

×.

Query by Observation

Measurements are from **same** sample instance Entity=Kelp AND Characteristic=DryMass AND Characteristic=WetMass

Query by Observation

File Edit Search Docum	entation Data Annotation V	Nindow Help		
Current profile: benriver (uid=leinfelder,o=NCEAS,dc=ecoint	ormatics,dc=org)		2010 (d)	Macinte
Change profile: benriv	Wel	come to Morpho	p!	
Network Status: NOT Log	ged In will be able to access only			-
Login to network using cur Password:	rent profile:			every
Work with your data:				
Create a new data packa Open an existing data p	ige			
Search for an existing d	ata package			
	100 200 - 100 100 200 - 100 100 200 - 100	and the second s		
and a second second				

🐮 Q

Load Ontology

Load custom OBOE-compatible ontology

- Santa Barbara Coastal LTER ontology
- Plant Trait Ontology
- Others

Load and Use Multiple Ontologies

🖲 Stop Recording 🕘 📕 🕴 🛜 🖌 💽 (Charged) 💽 Thu 11:04 PM 💄 QuickTime Player File Edit View Share Window Help Q Morpho File Edit Search Documentation Data Annotation Window Help V C C D D 4 000 Annotation Search - Specify one or more search criteria. Current profile: Macintosh HD (uid=leinfelder,o=N + Add Entity, Characteristic, Standard, and Protocol classes individually, or a single Measurement type can be selected as a template. <+> Add Context criteria. Change profile [+] Add Grouped criteria. 'Match All' requires that all criteria in the group are met. Create a nev V Local Network Status: Network If you do not choo "public" network file Match All Characteristic \$ is \$ -Login to netwo Characteristic everything . Password: <+> [+] Work with your d Create a ne Open an ex Search for OK Cancel - 0 6 **b** .

Future Directions

- Continue building corpus of data to annotate
- Refine "design patterns" for observation compliant domain ontologies
- Align/integrate ontologies at common points
 - Mass, units
- Iterate design for annotation interface
- Stronger inferencing: measurement types, transitivity along properties (e.g., partonomy), data "value-based" querying

Semi-automated aggregation, integration

Acknowledgements

SONet: Mark Schildhauer (NCEAS), Luis Bermudez (SURA), Flip Dibner (OGCii), Shawn Bowers (Gonzaga), Huiping Cao (NCEAS/ NMSU, Corinna Gries (UWisc Madison), Matt Jones (NCEAS), Deborah McGuinness (RPI)

Semtools: Matt Jones (NCEAS), Shawn Bowers (Gonzaga), Bertram Ludaescher (UC Davis), Margaret O'Brien (UCSB), Mark Schildhauer (NCEAS)

Thanks: Marie-Angelique LaPorte (CEFE/CNRS- Montpellier), Farshid Ahrestani (TraitNet/Columbia), Daniel Bunker (TraitNet, NJIT)

Acknowledgements

This material is based upon work supported by: The National Science Foundation under Grant Numbers 9980154, 9904777, 0131178, 9905838, 0129792, 0225674, 0225676, 0743429, 0733849, 0753144, 0630033.