

Ontologies and Linked Open Data in the LifeWatch Greece Research Infrastructure

Carlo Allocca, Martin Doerr, Chryssoula Bekiari, Nicolas Bailly, Nikos Minadakis, Yannis Marketakis, Dimitra Mavraki, Stamatina Nikolopoulou, Christos Arvanitidis

Presenter: Yannis Marketakis







- Problem, Approach and Objectives
- Ontologies and Linked Open Data
- On Generating URIs
- MarineTLO
- CIDOC CRM and CRM SCI
- Modeling the Sampling Process
- Concluding Remarks



Problem, Approach and Objective

- <u>The problem:</u>
 - Data providers usually use their own schemata to describe their data
 - Too much heterogeneity
 - The process is rarely well-documented
- Our Approach:
 - Provide a conceptual framework that will allow modeling biodiversity data and the biological observation processes
 - Through the use of ontologies and publishing data as Linked Data

• The Objectives:

- Enhance the formal representation of biological data and biological observation process
- Enable data integration
- Make implicit knowledge explicit



 An Ontology is the formal naming and definition of types, properties and interrelationships of the entities of a domain of discourse

- People understand semantics, machines don't

- Linked Open Data
 - Use Unique Resource Identifiers (URIs) as names for things
 - Use HTTP URIs so that people can look up those names
 - Provide useful information about what a name identifies when it's looked up
 - Refer to other things using their HTTP URI-based names





On Generating URIs

- Requirements
 - [R1] Avoid un-named (blank) nodes
 - Always assign URIs to resources
 - [R2] Avoid Collisions
 - Use a combination of attributes to minimize the risk of collisions (depending on the resource)
 - [R3] Avoid Duplications
 - Do not create URIs for resources that already exist
- An example for Thunnus albacares
 - <Base>/<EntityCode>/<combination_of_attributes>
 - http://www.lifewatchgreece.eu/Dataset/Species/Thunnus_Albacares



MarineTLO

- MarineTLO aims at being a global core model that
 - provides a common, agreed-upon and understanding of the concepts and relationships holding in the marine domain to enable knowledge sharing, information exchanging and integration between heterogeneous sources
 - covers with suitable abstractions the marine and the terrestrial domain to enable the most fundamental queries,
 - can be extended to any level of detail on demand, and
 - allows data originating from distinct sources to be adequately mapped and integrated
- The Latest version (V.4)
 - 127 Classes
 - 81 Properties
 - (<u>http://www.ics.forth.gr/isl/MarineTLO/</u>)



CIDOC CRM and CRMsci

- **CIDOC CRM** is a core ontology which is intended to promote a shared understanding of cultural heritage information.
- CRMsci is a formal ontology indented to be used as a global schema for integrating metadata about scientific observation, measurements and processed data in descriptive and empirical sciences such as biodiversity, geology, geography, archaeology, and others.
- It uses and extends CIDOC CRM (ISO 21127) <u>http://www.ics.forth.gr/isl/CRMext</u>





Modeling the Sampling Process





















Modeling the Sampling Process



October 8, 2015

ICZEGAR'2015 - Heraklion



Concluding Remarks

- The proposed approach allows:
 - Enhancing the formal representation of biological data and the biological processes
 - Expressing implicit knowledge, explicitly
 - Enabling the data integration from heterogeneous sources
 - Interlinking biological information
 - Answering complex queries that cannot be formulated, or answered from the particular data sources



Thank You !

October 8, 2015

ICZEGAR'2015 - Heraklion



Supplementary material





CIDOC CRM and CRMsci – cont'd





4	id	1Channel-1474		
5	language	EN		
6				Generic into about the datase
7	datasetID	4		
8	institutionCode	Hellenic Centre for Marine Research		
9	collectionCode	GR-Benthic communities in Amvrakikos Wetlands: Mazoma. Tsopeli.Tsoukalio. Rodia and Logarou lagoons-4		
10	basisOfRecord	Human observation		
11	informationWithheld	taxon remarks represents the original species name as written to the original dataset		
12	occurrenceID	1Channel-1474		
13	catalogNumber	GR-Amvrakikos Benthic communities-1Channel-1474		
14	samplingProtocol	box corer. 0.03 square meters		
15	eventDate	9/13/10		
16	habitat	coastal lagoon		
17	fieldNumber	1Channel		
18	locationID	Channel		
19	higherGeographyID	http://www.marineregions.org/gazetteer.php?p=details&id=3665		
20	higherGeography	Mediterranean Sea - Eastern Basin-Ionian Sea-Amvrakikos Kolpos		
21	waterBody	Ionian Sea		
22	country	Greece EEZ		
23	stateProvince	Epirus	ſ	- Geo-Into
24	locality	Amvrakikos Kolpos - Ro	dia lagoon	
25	maximumDepthInMeters	3		
26	verbatimCoordinates	Channel: 39B°04'36.4" N; 20B°48'40.5"		
27	decimalLatitude	39.07677778		
28	decimalLongitude	20.81125		
29	scientificNameID	urn:lsid:marinespecies.org:taxname:141438		
30	scientificName	Abra segmentum	Ī	
31	kingdom	Animalia		
32	phylum			
33	class	Bivalvia		
34	order	Veneroida		
35	family	Semelidae Worms Taxonomic info		
36	genus	Abra		
37	subgenus			
38	specificErsthet 15	segmentum		23
39	scientificNameAuthorship	(RF©cluz, 1843)		
40	taxonRemarks	Abra segmentum		