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
Outline

• 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

• **The problem:**
  – 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
Ontologies and Linked Open Data

- 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 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
- (http://www.ics.forth.gr/isl/MarineTLO/)
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) [http://www.ics.forth.gr/isl/CRMext](http://www.ics.forth.gr/isl/CRMext)
Modeling the Sampling Process

Standard Dataset

Dataset

refers to

Original (detailed) Dataset

Sampling Activity

Sampling Activity

Sampling Activity
Modeling the Sampling Process

Step 1: the protocol tells me to measure the position, ...ok!

Position Measurement

Sampling Activity

consists of

observed dimension

Dimension

Dimension
Modeling the Sampling Process

I need to take samples at the depths of (1m, 5m, 10m)

following the protocol...

Sampling Activity
I need to take samples at the depths of (1m, 5m, 10m) following the protocol...

Physical Object

Sampling Activity

1 m

s-1-A

s-1-B

s-1-C

5 m

s-5-A

s-5-B

s-5-C

10 m

s-10-A

s-10-B

s-10-C
Modeling the Sampling Process

I need to take samples at the depths of (1m, 5m, 10m)

following the protocol...

Physical Object

Sampling Activity

consists of

Encounter Event

has found object
I need to take samples at the depths of (1m, 5m, 10m) following the protocol...

Physical Object

1 m

s-1-A
s-1-B
s-1-C

5 m

s-5-A
s-5-B
s-5-C

10 m

s-10-A
s-10-B

Sampling Activity

consists of

Encounter Event

has found object

Physical Object

Physical Object

Physical Object

Physical Object

Physical Object

Physical Object
I need to take samples at the depths of (1m, 5m, 10m) following the protocol...
I need to take samples at the depths of (1m, 5m, 10m) following the protocol...
Modeling the Sampling Process

Dataset

"linked" data

"flat" data

Sampling Activity

Position Measurement

Physical Object

Measurement

Dimension

observed dimension

has found object

measured

consists of

refers to

Consists of

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 !
Supplementary material
MarineTLO – cont’d

TLO Entity

Observable Entity

Event

Activity

Temporal Phenomenon

Ecosystem

Time Span

Dimension

Place

Declarative Place

Persistent Item

Thing

Man-made Thing

Physical Thing

Actor

Organization

Group

Software Execution

Attribute Assignment

Capture Activity

Observation

Identifier Assignment

Global Statistic Landing

Name Use Activity

Proposition

Appellation

Measurement Unit

Design or Procedure

Statistic Indicator

Conceptual Object

Digital Device

Physical Man-made Thing

Ecosystem Environment

Water Area

Abiotic Element

Physical Object

Capture

Biological Object

Collection

Man-made Object
CIDOC CRM and CRMsci – cont’d
<table>
<thead>
<tr>
<th>id</th>
<th>1Channel-1474</th>
</tr>
</thead>
<tbody>
<tr>
<td>language</td>
<td>EN</td>
</tr>
<tr>
<td>datasetID</td>
<td>4</td>
</tr>
<tr>
<td>institutionCode</td>
<td>Hellenic Centre for Marine Research</td>
</tr>
<tr>
<td>collectionCode</td>
<td>GR-Benthic communities in Amvrakikos Wetlands: Mazama, Tsopeli, Tsoukalio, Rodia and Logarou lagoons-4</td>
</tr>
<tr>
<td>basisOfRecord</td>
<td>Human observation</td>
</tr>
<tr>
<td>informationWithheld</td>
<td>taxon remarks represents the original species name as written to the original dataset</td>
</tr>
<tr>
<td>occurrenceID</td>
<td>1Channel-1474</td>
</tr>
<tr>
<td>catalogNumber</td>
<td>GR-Amvrakikos Benthic communities-1Channel-1474</td>
</tr>
<tr>
<td>samplingProtocol</td>
<td>box corer. 0.03 square meters</td>
</tr>
<tr>
<td>eventDate</td>
<td>9/13/10</td>
</tr>
<tr>
<td>habitat</td>
<td>coastal lagoon</td>
</tr>
<tr>
<td>fieldNumber</td>
<td>1Channel</td>
</tr>
<tr>
<td>locationID</td>
<td>Channel</td>
</tr>
<tr>
<td>higherGeography</td>
<td>Mediterranean Sea - Eastern Basin-Ionian Sea-Amvrakikos Kolpos</td>
</tr>
<tr>
<td>waterBody</td>
<td>Ionian Sea</td>
</tr>
<tr>
<td>country</td>
<td>Greece EEZ</td>
</tr>
<tr>
<td>stateProvince</td>
<td>Epirus</td>
</tr>
<tr>
<td>locality</td>
<td>Amvrakikos Kolpos - Rodia lagoon</td>
</tr>
<tr>
<td>maximumDepthInMeters</td>
<td>3</td>
</tr>
<tr>
<td>verbatimCoordinates</td>
<td>Channel: 39°04′36.4″ N; 20°48′40.5″</td>
</tr>
<tr>
<td>decimalLatitude</td>
<td>39.07677778</td>
</tr>
<tr>
<td>decimalLongitude</td>
<td>20.81125</td>
</tr>
<tr>
<td>scientificNameID</td>
<td>urn:lsid:marinespecies.org:taxname:141438</td>
</tr>
<tr>
<td>scientificName</td>
<td>Abra segmentum</td>
</tr>
<tr>
<td>kingdom</td>
<td>Animalia</td>
</tr>
<tr>
<td>phylum</td>
<td>Bivalvia</td>
</tr>
<tr>
<td>class</td>
<td>Veneroida</td>
</tr>
<tr>
<td>order</td>
<td>Semelidae</td>
</tr>
<tr>
<td>genus</td>
<td>Abra</td>
</tr>
<tr>
<td>subgenus</td>
<td>Abra segmentum</td>
</tr>
<tr>
<td>specificEpithet</td>
<td>segmentum</td>
</tr>
<tr>
<td>scientificNameAuthorship</td>
<td>(RF©cluz. 1843)</td>
</tr>
<tr>
<td>taxonRemarks</td>
<td>Abra segmentum</td>
</tr>
</tbody>
</table>