Knowledge Base for Global Registry of Stocks and Fisheries

Nikos Minadakis, FORTH-ICS

minadakn@ics.forth.gr

BlueBRIDGE TWG Meeting
1-2 March 2016
Rome, Italy

BlueBRIDGE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 675680

Outline

1. Knowledge Base Construction Process
2. Top Level Ontologies and Models
3. iMarine Application
4. Blue Bridge Knowledge Base Progress
5. Examples

Constructing a KB Process

Collection of Functional Requirements

Collection & Analysis of Domain Datasets

Collection of the Competency Queries

Selection of the Semantic Models

Data Transformation, Ingestion and Eval

Cr

De

Extension of the Semantic Models

Ontologies

- ***Ontology*** is a ***formal naming*** and ***definition*** of:
  a) *the types*
  b) *the properties*
  c) *the interrelationships*

  of the ***entities*** for a particular ***domain of discourse***.

- Some of the ***advantages*** of using ontologies are:
  - ***Machines*** understand the ***meaning*** of information
  - Data ***linking***
  - Consistent ***global identifiers***
  - Information ***integration*** / ***aggregation***
  - Efficient ***data discovery*** methods

Ontologies & Semantic Models

- **MarineTLO**:
  - Global ***core model*** for the marine domain
  - Can be applied to the ***terrestrial*** domain
  - Enables ***information exchanging*** and ***integration*** between heterogeneous sources.
  - Easy ***extendibility***
  - Has been used in ***iMarine*** and ***LifeWatch Greece***

- **CIDOC – CRM & Extensions:**
  - ***Describes concepts and relationships*** used in cultural heritage domain in a philosophical way.
  - Can be applied effectively to a variety of domains such us ***biodiversity, geology, etc***.
  - Official standard ***ISO 21127:2006***.
  - Extensions:
    - ***CRMsci, CRMdig, CRMgeo, CRMinf***

Conceptual Modeling Example

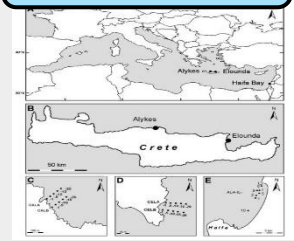
S40 Encounter Event
urn:catalog:IOL:POLY:Sphaerosyllis-levantina-ALA-IL-7-Oct.2009

P4 has timespan → **E52 Time-Span**
11/10/2009

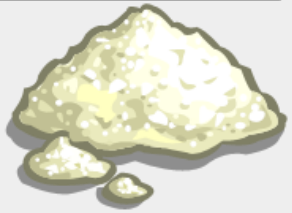
P14 carried out by → **E21 Person**
S. Faulwetter

P7 took place at → **BC14 Ecosystem Environment**
Haifa Bay Ecosystem No

P10 falls within → **E53 Place**
Alykes

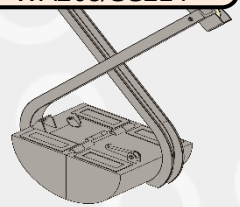


P2 has type → **BT7 Ecosystem Type**
sandy - muddy sediments



O32 has found object → **BC38 Biotic Element**
Sphaero-levantina-003

P125 used object of type → **BT11 Equipment Type**
WA265/SS214



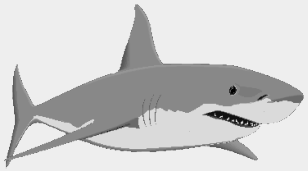
P127 has broader term → **BT11 Equipment Type**
Van Veen Grab

iMarine Warehouse Contents

5.5M Triples

53,797 Species

2,054 Predators

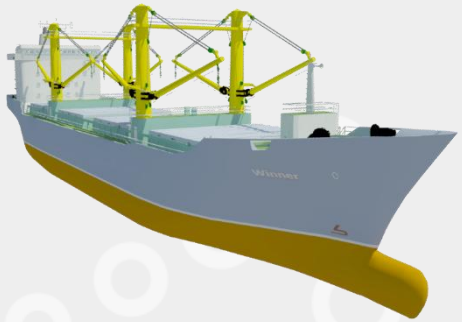


53,797 Scientific Names
12,752 Authorships
155,510 Common Names

853 Ecosystems
294 Countries

47 Water Areas
160 EEZ

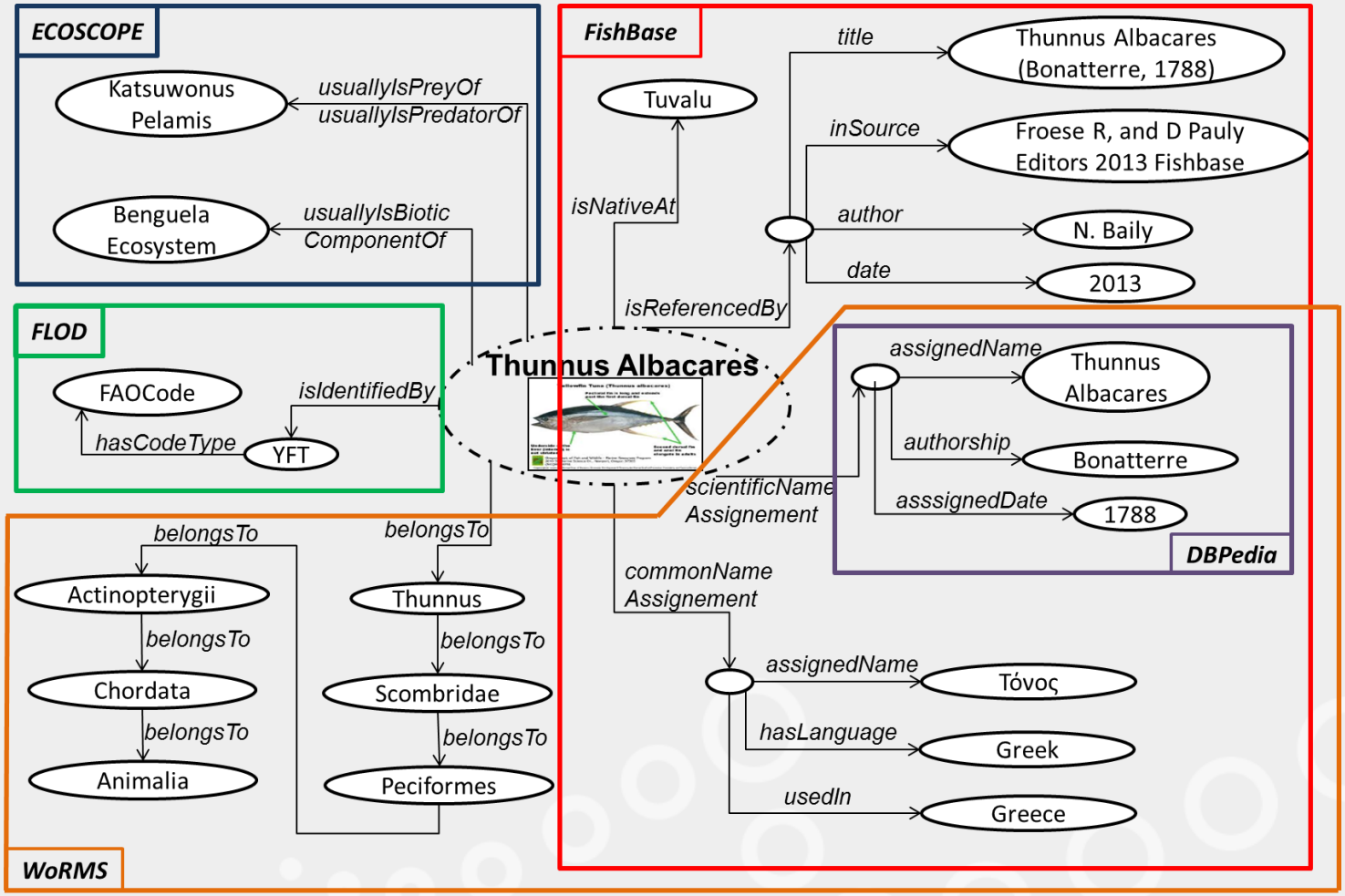
90 Vessel Types
112 Gear Types



13,131 Bibliography resourc
3,206 Statistic Indicators

iMarine Integration Example

Find the scientific name of the fish that is usually called "τόνος" in Greece, along its preys and predators, and the area they are native at.

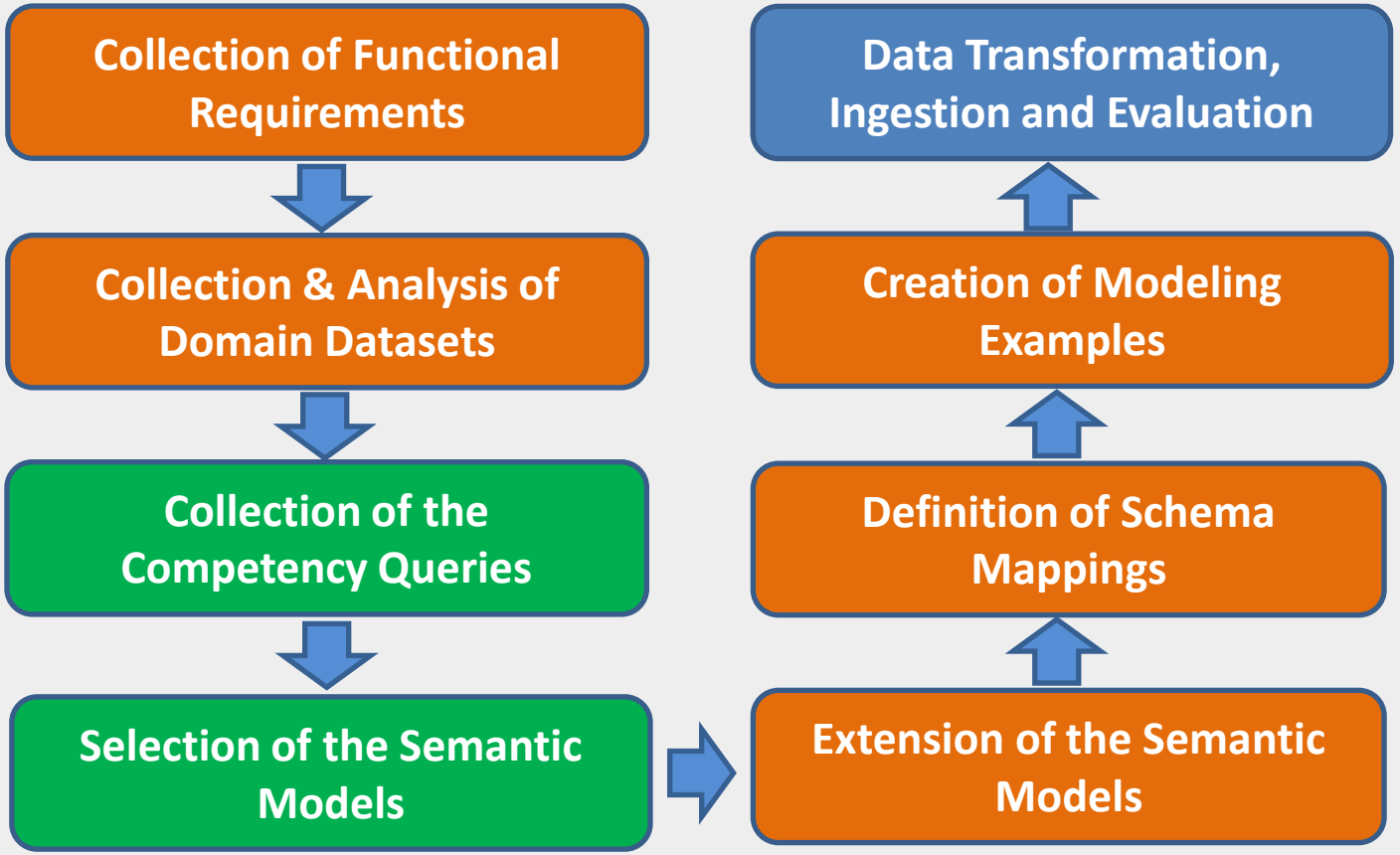


BB GRSF Knowledge Base Goals

Construction of a ***global knowledge base*** for the Global Record of Stocks and Fisheries (GRSF):

- ***Integrating*** heterogeneous data
- ***Publishing*** data according to LOD principles
- ***Connecting*** data
- ***Exposing*** stocks and fisheries data
- ***Enabling complex query*** answering
- ***Building tools*** and ***web applications*** that will exploit the knowledge base

Knowledge Base Progress



in progress

completed

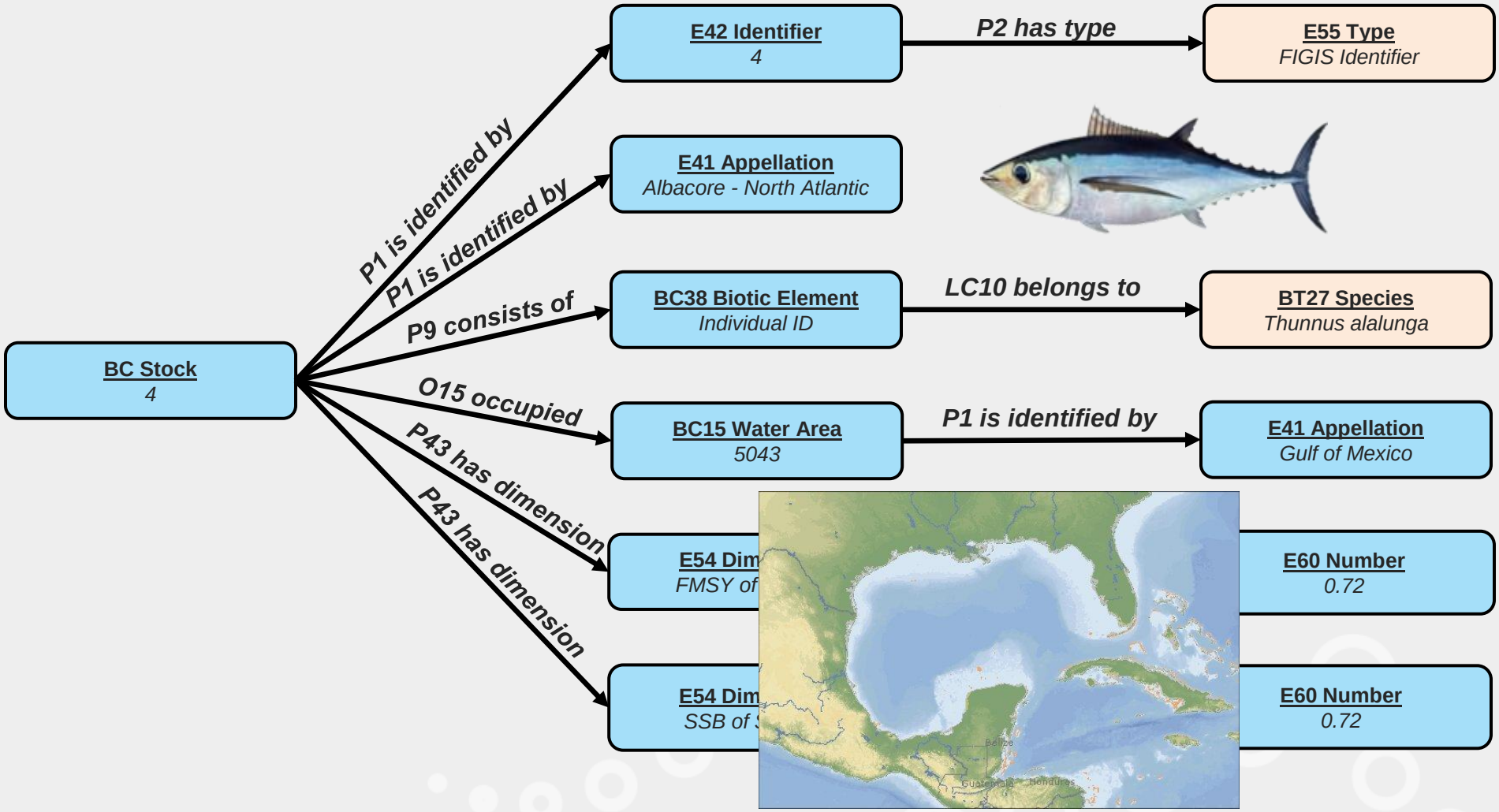
pending

Minimum Data Requirements

Concept (Stocks)	Covered
Stock/Assessment unit ID	Yes (Validate)
Stock/Assessment unit Name	Yes
Species	Yes
Fishing area	Yes
Exploiting fishery	Yes (Validate)
Management Unit	Not Yet
Assessment Methods	Yes (Validate)
Scientific Advice	Yes (Validate)
State of Marine Resources	Not Yet
Exploitation Rate	Yes (Validate)
Abundance Level	Yes (Validate)
DB Source	Not Yet
Source of Information	Yes
Ownership	Yes

Concept (Fisheries)	Covered
Fishery ID	Yes (Validate)
Fishery Name	Yes
Species	Yes
Fishing area	Yes
Exploiting stocks	Yes (Validate)
Management unit	Not Yet
Prod system type	Not Yet
Flag state	Yes (Validate)
Fishing gear	Yes
Annual Catch	Yes (Validate)
DB Source	Not Yet
Source of Information	Yes
Ownership	Yes

Stocks Modeling



Next Steps

1. **Continue data analysis and modeling and finalize mappings (include FishSource)**
   - FIRMS
   - RAM Legacy Database
   - FishSource (waiting resources from the FishSource team)

2. **Start exporting data from FIRMS w.r.t. to MarineTLO**

3. **Integrate data to MarineTLO-based warehouse**

4. **MatWare & x3ml Engine integration**

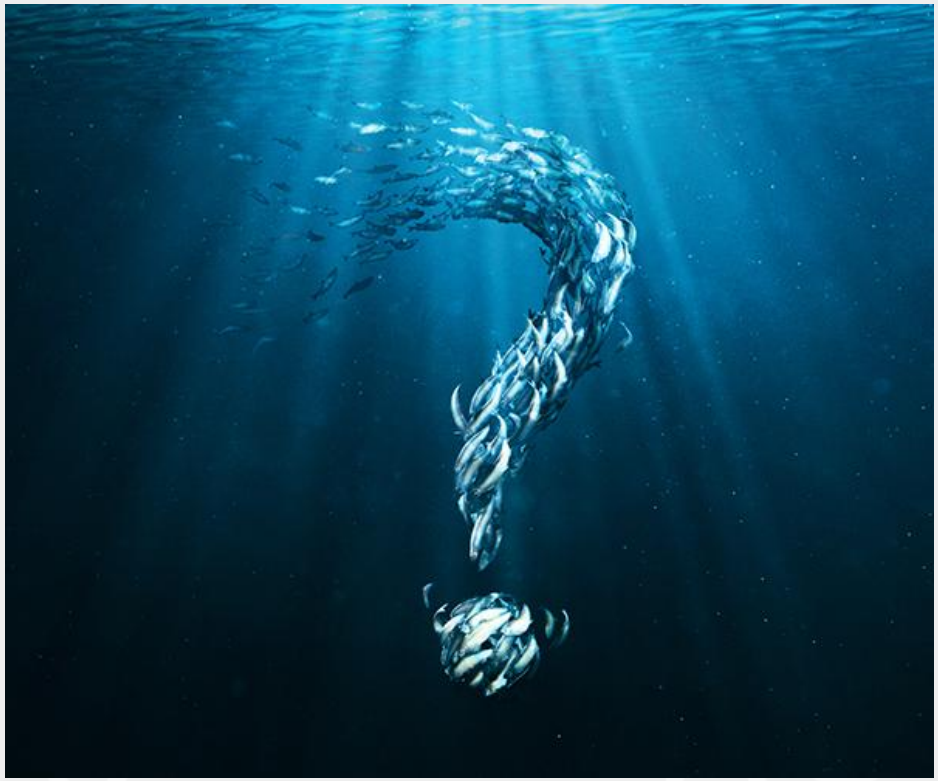
5. **Continue the activities regarding 3D visualizations**

References

- MarineTLO
  - http://wiki.i-marine.eu/index.php/Top_Level_Ontology
  - http://wiki.i-marine.eu/index.php/MarineTLO-based_warehouse
  - http://www.ics.forth.gr/isl/MarineTLO/
- MarineTLO-based warehouse
  - https://i-marine.d4science.org/
    - Though the VREs BiodiversityLab, MarineSearch, iSearch
- MatWare: *Automates the construction and evaluation of the connectivity of a semantic warehouse*
  - http://www.ics.forth.gr/isl/MatWare/
- Connectivity Metrics & VoID extension: A set of metrics to assess the connectivity of sources and the quality of the constructed Knowledge Base
  - http://www.ics.forth.gr/isl/MatWare/files/tzitzik2014_connectivity_LWDM.pdf
  - http://www.ics.forth.gr/isl/VoIDWarehouse
- CIDOC and CRM Family models
  - http://www.cidoc-crm.org/
  - http://www.ics.forth.gr/isl/CRMext
- X3ml Engine: *Handles the URI generation and the data transformation steps of the data provision and aggregation process*
  - https://github.com/delving/x3ml

BlueBRIDGE

THANK YOU!

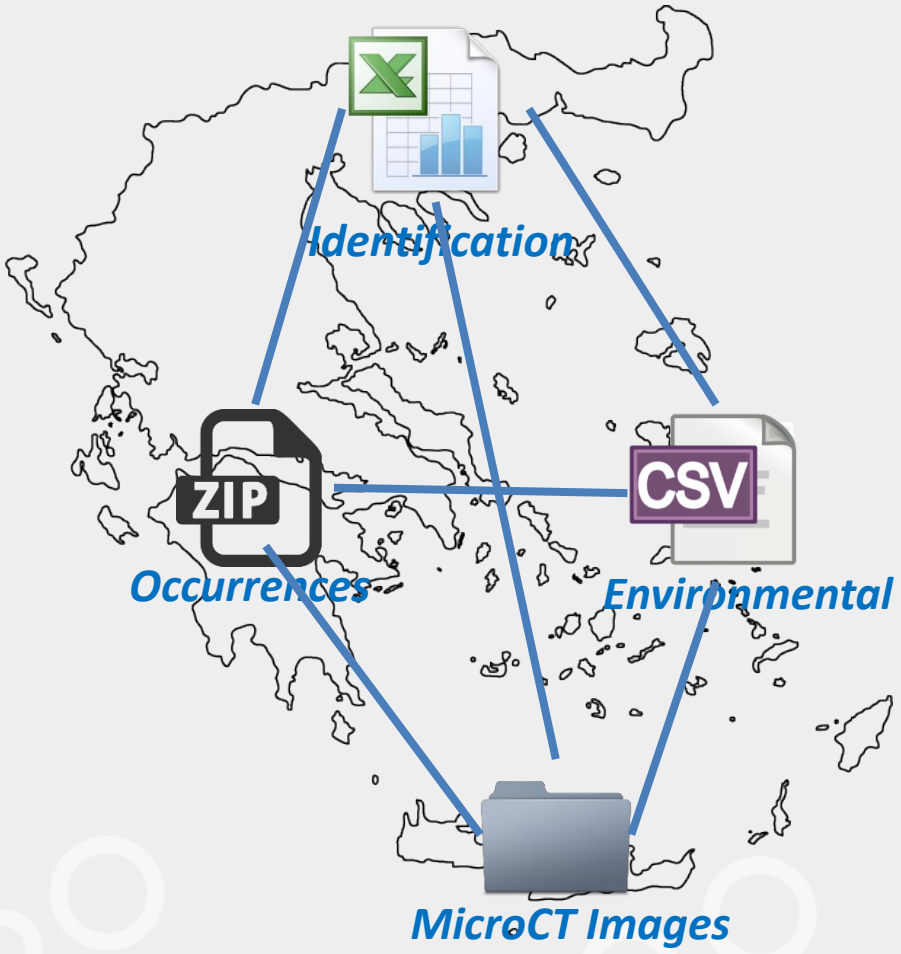


FORTH
Institute of
Computer Science

APPENDIX

LifeWatch Greece

- ***supporting cataloguing and publishing of*** all the relevant meta-data information of the Greek biodiversity domain.
- ***integrating data from heterogeneous sources*** by supporting the definitions of appropriate models.
- ***efficiently discovering biodiversity data*** of interest and enable the answering of complex queries that could not be answered from the individual sources.



LifeWatch Data Discovery

A user searches for occurrences of species that took place in the Mediterranean sea.

Occurrences Information Found

Row	Species	Country	Date	Related Dataset	
1	Apomatus globifer	Libya	1890	View dataset	More info
2	Protula tubularia	Greece	1891	View dataset	More info
3	Protula marioni	Italy	1892	View dataset	More info
4	Serpula vermicularis	Italy	1892	View dataset	More info
5	Serpula vermicularis	Greece	1890	View dataset	More info
6	Laonome salmacidis	Turkey	1892	View dataset	More info
7	Polycirrus aurantiacus	Greece	1891	View dataset	More info
8	Hydroides norvegica	Greece	1890	View dataset	More info
9	Melinna adriatica	Israeli Exclusive Economic Zone	1892	View dataset	More info
10	Placostegus tridentatus	Turkey	1892	View dataset	More info

LifeWatch Data Discovery

Then user clicks on the individual that was found during the event

7	Polycirrus aurantiacus	Greece	1891	View dataset	More info

Dataset Title:	Pola_Expedition_Polychaeta
Occurrence Event ID:	http://www.lifewatchgreece.eu/entity/encounterEvent/pola_expedition_polychaeta_15
Individual ID:	polycirrus_aurantiacus_1891_15
Actor:	Panagiotis Damianidis
Place:	east of Kythira
Water Area:	Aegean Sea
Habitat:	Mud with rocks
Equipment Type:	
Ecosystem:	
Description:	
Station URI:	http://www.lifewatchgreece.eu/entity/place/stationpola_expedition_polychaeta_15
Station Notes:	maximumDepth:415,minimumDepth:415
Bibliographic Citation:	
Sampling Protocol:	
Coordinates:	36.0917,23.1583

LifeWatch Data Discovery

Then user clicks on the transformation event of the individual

Info about node: polycirrus_aurantiacus_1891_15

Row	Relation	Object	Object Type
1	type	BC38_Biotic_Element	-
2	LC10_belongs_to	Polycirrus aurantiacus	BT27_Species

Row	Subject	Subject Type	Relation
1	http://www.lifewatchgreece.eu/entity/identificationEvent/pola_expedition_polychaetapo	E17_Type_Assignment	P41_classified
2	http://www.lifewatchgreece.eu/entity/encounterEvent/pola_expedition_polychaeta_15	S19_Encounter_Event	O32_has_found_object
3	http://www.lifewatchgreece.eu/entity/transformationEvent/transfomationeventid7	E81_Transformation	P124_transformed

LifeWatch Data Discovery

Then user clicks on specimen that was created from the transformation event

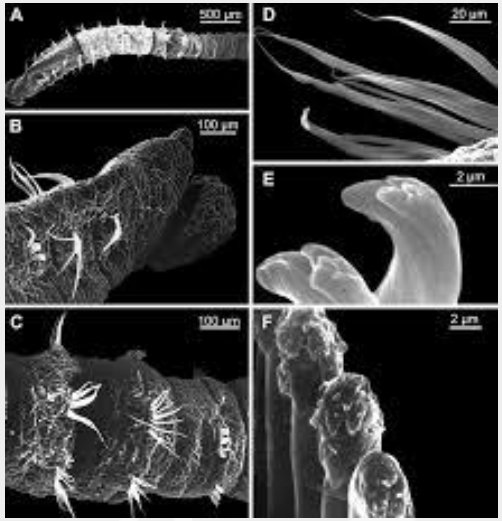
Info about node: http://www.lifewatchgreece.eu/entity/transformationEvent/transfomationeventid7

Row	Relation	Object	Object Type
1	type	E81_Transformation	-
2	P124_transformed	polycirrus_aurantiacus_1891_15	BC38_Biotic_Element
3	P123_resulted_in	mCT-00038	BC53_Specimen
4	P4_has_timespan	2011-09-26	-
5	P14_carried_out_by	Matina Nikolopoulou	E21_Person
6	P33_used_specific_technique	preserved in ethanol	

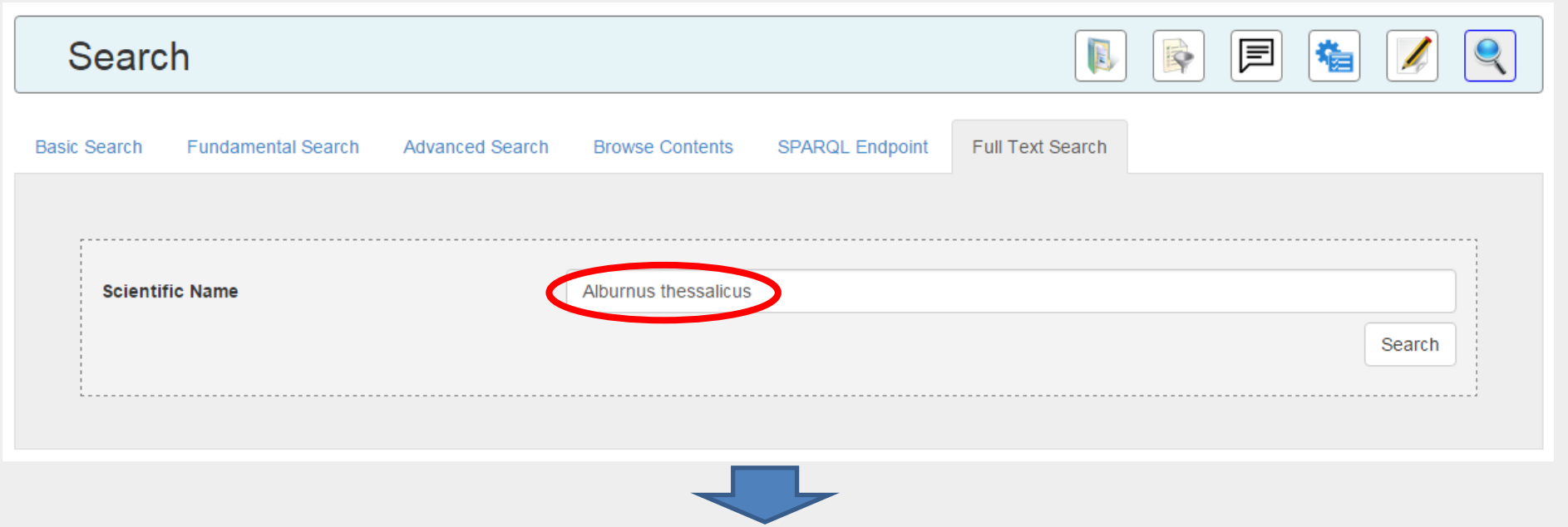
LifeWatch Data Discovery

Row	Specimen	Device	Enhancement Method	Related Dataset	
1	mCT-00038	SkyScan 1172	none	View dataset	More info

Dataset Title: Pola_Expedition_Polychaeta
Scanning ID: http://www.lifewatchgreece.eu/entity/digitizationProcess/scan-00039
Product: scan-00039.zip
Actor: Sarah Faulwetter
Date: 7/9/2012



LifeWatch Data Discovery



The ***Alburnus thessalicus*** species belongs to the ***Alburnus*** genus, ***Cyprinidae*** family, ***Cypriniformes*** order , ***Actinopterygii*** class, ***Chordata*** phylum and ***Animalia*** kingdom. ***Alburnus thessalicus*** was discovered by ***Alexander I. Stephanidis*** in 1950. Individuals that belong to this species have been occurred in ***Greece***, ***Italy***, ***Spain*** etc. Individual ***alburnus_thessalicus_22*** was transformed into ***alb_thes_22_specimen*** by ***Sara Faulwetter*** in 2015. ***alb_thes_22_specimen*** was scanned by ***Niki Keklikoglou*** in 2014 using a ***SkyScan*** MicroCT tomograph and results to the ***alb_thes_scans.zip*** dataset.