

***Deliverable 2.2: Metadata Catalogue***

|  |  |
| --- | --- |
| Document Title | Deliverable 2.2: Metadata Catalogue |
|  |  |
| Due date | 31/10/2015 |
| Date of submission |  |
|  |  |
| Lead beneficiary | Hellenic Centre for Marine Research |
| Responsible | FORTH/ICS/ISL |
| Dissemination Level\*1 | RE |
| Nature | Deliverable |
| Status of the Document | 1st Draft |
| Version | 1.0 |
|  |  |
|  |  |

|  |  |
| --- | --- |
| **\*1 Dissemination Level** |  |
| **PU** | Public |
| **PP** | Restricted to other programme participants (including Commission services) |
| **RE** | Restricted to a group specified by the consortium (including Commission services) |
| **CO** | Confidential, only for members of the consortium (including Commission services) |

|  |  |  |  |
| --- | --- | --- | --- |
| **Editor(s):** | **Name** | **Organisation** | **Email** |
|  | Chrysoula Bekiari | FORTH | bekiari@ics.forth.gr |
|  | Martin Doerr | FORTH | martin@ics.forth.gr |
|  | Nikos Minadakis | FORTH | minadakn@ics.forth.gr |
|  | Athina Kritsotaki | FORTH | athinak@ics.forth.gr |
|  | Vasiliki Tsimpida | FORTH | vtsimpida1112@gmail.com |
|  | Yannis Marketakis | FORTH | marketak@ics.forth.gr |
|  |  |  |  |
| **Contributor(s):** | Christos Arvanitidis | HCMR | arvanitidis@hcmr.gr |
|  | Eva Chatzinikolaou | HCMR | evachatz@hcmr.gr |
|  | Nicolas Bailly | HCMR | nbailly@hcmr.gr |
|  | Sarah Faulwetter | HCMR | sarifa@hcmr.gr |
|  | Niki Keklikoglou | HCMR | keklikoglou@hcmr.gr |
|  | Jacques Lagnel | HCMR | lagnel@hcmr.gr |
|  | Dimitra Mavraki | HCMR | dmavraki@hcmr.gr |
|  | Matina Nikolopoulou | HCMR | snikolo@hcmr.gr |
|  | Anastasis Oulas | HCMR | oulas@hcmr.gr |
|  | Evangelos Pafilis | HCMR | pafilis@hcmr.gr |
|  | Emmanouela Panteri | HCMR | emmipan@hcmr.gr |
|  | Christina Pavloudi | HCMR | cpavloud@hcmr.gr |

Table of Contents

[Metadata Catalogue Description 4](#_Toc435550167)

[Metadata Categories 6](#_Toc435550168)

[Occurrence Event Metadata 6](#_Toc435550169)

[Identification Event Metadata 8](#_Toc435550170)

[Naming and Taxonomic Metadata 9](#_Toc435550171)

[Micro-CT processes metadata 11](#_Toc435550172)

[Environmental Measurement Metadata 16](#_Toc435550173)

[Specimen Metadata 17](#_Toc435550174)

[Specimen Collection Metadata 19](#_Toc435550175)

[Data Collection Metadata 20](#_Toc435550176)

[Genetic Events Metadata 22](#_Toc435550177)

[Measurement Event Metadata 24](#_Toc435550178)

[Morphological Characteristics Metadata 25](#_Toc435550179)

[Statistical Metadata 27](#_Toc435550180)

[Deliverable 2.2 Report 29](#_Toc435550181)

# Metadata Catalogue Description

Deliverable 2.2 Metadata Catalogue is part of the ***Task 2.2: Data Providers Services Platform***. This metadata catalogue contains all the metadata information used by the data providers, listed. The catalogue was continuously evolving and expanded according to the new types of metadata and datasets that were collected by the community, until it took its final form in October 2015.

During the first LifeWatchGreece Data Management workshop - which gathered Data Managers representing Research Institutes and Universities from all over Greece - an extended and detailed discussion took place regarding “data collection” and their entry into the LWG repository. One of the tasks discussed and approved was that each Data Manager shall prepare by the end of September 2014, a list of datasets which are going to be stored at LWG RI. To this extent a Data Policy and Sharing Agreement legal document was prepared, highlighting the copyright issues, as well as the duties and rights, both of data providers and LifeWatchGreece RI.

As an outcome of this procedure, a significant number of metadata, datasets and competency queries was collected by the Greek biodiversity community and analyzed in order to conclude on which metadata is mandatory to keep in the national catalogue and to define the main metadata categories.

Specifically for occurrences and identifications, potential queries on the datasets from HCMR and NHMC were collected, the format of these data sets being mainly \*.csv, and \*.sql databases. These competency queries are:

1. *Which species have been originally described from Greece?*
2. *Which species have been originally described from the Mediterranean / Greece / Aegean Sea and afterwards have never been recorded again?*
3. *What is the recorded depth range for each polychaete species found so far in Greek waters?*
4. *Find all taxonomic names and their associated authorities that have ever been placed in genus X (even if now they are in another genus), preferably including the bibliographic reference of the original description, and deliver also their geographic distribution, including information on the locality, habitat and depth / elevation of the type specimen for each of those species*

For MicroCT scanning, datasets and metadata from HCMR were collected and the format of them was mainly \*.xls, \*.txt, \*.png, and \*.tiff. Some indicative competency queries are the following:

1. *Find all scans depicting marine species*
2. *Find all scans depicting polychaetes*
3. *Find all scans depicting specimens of 5mm length*
4. *Find all scans depicting specimens stained with Iodine*

For genetics, datasets from HCMR were collected and the format of them was mainly \*.xls, \*.fasta and \*.sff files. Some indicative competency queries are the following:

1. *How many/Which samples have been sequenced with e.g. Illumina HiSeq?*
2. *How many/Which samples have been analyzed with e.g. Qiime?*

For measurements and morphometrics, datasets from HCMR were collected and the format of them was mainly xls and csv.

For environmental parameters, datasets from HCMR were collected and the format of them was mainly \*.xls. Some indicative competency queries are the following:

1. *How many stations have pH x ?*
2. *Which lagoon has salinity x psu ?*
3. *Find the species that were found in salinity x Psu*

# 

# Metadata Categories

The analysis of the Greek biodiversity community’s datasets and metadata led to the identification of fourteen (14) main metadata categories than can cover the full set of metadata that should be kept in the metadata catalogue. This categorization of metadata affects the metadata repository design, the data publishing workflows and the querying functions. The twelve categories are presented below, containing the required metadata that belong to each category alongside with a related description.[[1]](#footnote-1)

## Occurrence Event Metadata

Occurrence event is the event of finding an individual of a species at a particular place (water area, country etc) at a particular time. Below, we present the table with the related metadata, their definition, some examples related to metadata of the occurrence event and the mapping to CIDOC – CRM family models (Table 1).

**Metadata about Occurrence**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| occurrenceID | The ID of the occurrence event that took place. | “urn:catalog:IOL:POLY:Sphaerosyllis-levantina-ALA-IL-7-Oct.2009 “ | S19 Encounter Event.  P48 has preferred identifier: E42 Identifier |
| dataset ID | The ID of the dataset that contains records that refer to this occurrence event. | "d50024ac-5268-477e-8559-933779553b34"  "EasternMedSyllids" | S19 Encounter Event. P67 is referred to by:  BC21 Data Set |
| organismID | The ID of the species’ individual that was found during the occurrence event. | “urn:lsid:zoobank.org:act:9CEE8F90-9596-49F6-AA22-BB79C0E816D9” | S19 Encounter Event.  O19 has found object:  E18 Physical Thing |
| scientificName | The scientific name of species that the individual that was found belongs to. | "Odontosyllis fulgurans" | BT27 Species.  P1 is identified by:  E42 Identifier |
| recordedBy | The person that carried out the occurrence event. | "Sarah Faulwetter" | S19 Encounter Event.  P14 carried out by:  E39 Actor (BT9 Actor Type, BC8 Actor) |
| eventDate | The date of the occurrence event. | "22/10/2012" | S19 Encounter Event.  P4 has time-span:  E52 Time-span |
| locality | The verbatim location of the event that took place. | "AmvrakikosKolpos - Rodia lagoon" "AmvrakikosKolpos - Logarou lagoon"  "Alykes" "Elounda" | S19 Encounter Event.  O21 has found at:  E53 Place |
| country | The country that the event took place in. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names. | "Greece " "Israel" | S19 Encounter Event.  P7 took place at:  E53 Place |
| waterArea | The water area that the occurrence event took place in. Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names | "IonianSea" "MediterraneanSea | S19 Encounter Event. P7 took place at:  E53 Place.  P89 falls within: BC15 Water Area |
| habitat | The type of the ecosystem environment that the occurrence event took place. | "coastal lagoon" "rocky substrate” | BT7 Ecosystem Type |
| equipmentType | The type of the equipment that was used during the occurrence | "WA265/SS21" "Van Veen Grab" | S19 Encounter Event.  P125 used object of type:  BT11 Equipment Type |
| decimalLatitude | The geographic latitude of the place that the occurrence event took place | "390.767.777.778" "390.539.722.222"7 | S19 Encounter Event.  P7 took place at:  E53 Place.  P87 is identified by:  E47 Spatial Coordinates |
| decimalLongitude | The geographic longitude of the place that the occurrence event took place | "2.081.125" "20.913" | S19 Encounter Event. P7 took place at:  E53 Place.  P87 is identified by: E47 Spatial Coordinates |
| maximumDepthInMeters | The maximum depth that the occurrence event took place | "3" "0,5" | S19 Encounter Event.  P7 took place at: E53 Place.  P3 has note:  E62 String |
| minimumDepthInMeters | The minimum depth that the occurrence event took place | "20" "45" | S19 Encounter Event.  P7 took place at:  E53 Place.  P3 has note:  E62 String |
| samplingProtocol | The method or protocol that used during this event | "box corer. 0.03 square meters"  "doi: 10.4067/S0717-6538200300020000" | S19 Encounter Event . P33 used specific technique:  E29 Design or Procedure |
| bibliographicCitation | A bibliographic reference to occurrence event. | "Odontosyllis fulgurans (IMBG-NaGISA-CALB-20B\_07)" | S19 Encounter Event.. P67B is referred to by:  BC27 Publication |
| occurenceRemarks | Textual remarks/notes about the occurrence event. | “1902 Danish expendition occurrences” | S19 Encounter Event..  P3 has note:  E62 String |

*Table 1: Metadata about Occurrence Event*

## Identification Event Metadata

Identification event is the event of taxonomic determination of an organism. We present the table with the metadata, the definition, some examples related to identification event and the mapping to CIDOC – CRM family models (Table 2).

**Metadata about Identification**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| identificationID | The ID of the Identification event that took place | "20724" | E17 Type Assignment Event.  P1 is identified by:  E42 Identifier |
| organismID | The ID of the individual that was identified. | “Paraehlersia\_ferrugina-CALA-20C\_07” | BC38 Biotic Element.  P1 is identified by:  E42 Identifier  &  E17 Type Assignment Event. P41 Classified:  BC38 Biotic Element |
| identifiedBy | The person that carried out the identification event. | "Sarah Faulwetter" | E17 Type Assignment Event. P14 carried out by:  E39 Actor (BT9 Actor Type, BC8 Actor) |
| dateIdentified | The date of the identification event. | "17/9/2012" | E17 Type Assignment Event. P4 has time-span: BC2 Time Span |
| locality | The place of the identification event. | "Alykes" | E17 Type Assignment Event.  P7 took place at: E53 Place |
| scientificName | The scientific name of species that was assigned to the individual/specimen by the identification event taxon id | "Odontosyllis fulgurans" | E17 Type Assignment Event.  P42 assigned:  BT27 Species |
| identificationReferences | A list of references (publication, global unique identifier, URI) used in the Identification. | "SanMartín 2003" "doi:10.3897/zookeys.150.1877" | E17 Type Assignment Event.  P33 used specific object:  BC27 Publication |
| dataset ID | The ID of the dataset that contains records that refer to this identification event. | "d50024ac-5268-477e-8559-933779553b34"  "EasternMedSyllids" | E17 Type Assignment Event.  P67 is referred to by:  BC21 Data Set |

*Table 2: Metadata about Identification Event*

## Naming and Taxonomic Metadata

Taxonomy is an event that organisms are classified according to existing ordered systems which indicate natural relationships. The taxonomic organization of species is hierarchical. Species belongs to genus, genus belongs to family, and so on through order, class, phylum, and kingdom. In the following tables the information that we keep about the scientific name assignment events, the taxonomic classification and the common names are included.

**Metadata about Scientific Name Assignment Event**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM Family Models** |
| scientificNameAssignmentEventID | The ID of the event that assigned a scientific name to the taxon type | "urn:lsid:zoobank.org:act:9CEE8F90-9596-49F6-AA22-BB79C0E816D9" | E15 Identifier Assignment. P17 was modivated by: E83 Type Creation. P48 is identified by: E42 Identifier |
| scientificNameAuthorship | The person(s) that assigned a scientific name to the taxon type | "Sarah Faulwetter" | E15 Identifier Assignment  /E83 Type Creation.  P14 carried out by :  E39 Actor ( BT9 Actor Type, BC8 Actor) |
| namePublishedInYear | The date that a scientific name was assigned to the taxon type | "19/10/2012" | E83 Type Creation. P4 has time-span:  E52 Time Span |
| scientificName | The scientific name that was assigned to the taxon type | "Abrasegmentum" "Copepoda" | E15 Identifier Assignment. P141 assigned: E41 Appellation  (BT14 Appellation Type, BC30 Appellation) |
| nomenclaturalCode | The nomenclatural code (or codes in the case of an ambiregnal name) under which the scientificName is constructed. Recommended best practice is to use a controlled vocabulary. | "ICZN" | E15 Identifier Assignment. P33 used specific technique: E29 Design or Procedure |
| taxonID | The ID of the taxon that the scientific name was assigned to | “urn:lsid:zoobank.org:act:9CEE8F90-9596-49F6-AA22-BB79C0E816D9” | E15 Identifier Assignment.  P140 assigned to:  BT27 Species.  P1 is identified by:  E42 Identifier |
| dataset ID | The ID of the dataset that contains records that refer to this scientific name assignment event. | "d50024ac-5268-477e-8559-933779553b34"  "EasternMedSyllids" | E15 Identifier Assignment.  P67 is referred to by:  BC21 Data Set |

*Table 3*a*: Metadata about Scientific Name Assignment Event*

**Taxonomy Information**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **Marine TLO** |
| Species | The taxonID of the species | "urn:lsid:zoobank.org:act:9CEE8F90-9596-49F6-AA22-BB79C0E816D9" | BT27 Species |
| Genus | The genus in which the taxon is classified | "Capitella"1  "Caprella"2 | BT26 Genus |
| Family | The family in which the taxon is classified | "Gammaridae"3  "Spionidae"4 | BT24 Family |
| Order | The order in which the taxon is classified | "Amphipoda"5  "Spionida"6 | BT34 Order |
| Class | The class in which the taxon is classified | "Bivalvia"7  "Polychaeta"8 | BT22 Class |
| Phylum | The phylum in which the taxon is classified | "Annelida"9  "Nemertea"10 | BT19 Phylum |
| Kingdom | The kingdom in which the taxon is classified | "Animalia"11 | BT18 Kingdom |

*Table 3b: Taxon Ranks*

**Common Name Information**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM Family Models** |
| speciesID | The species that is identified by the common name. | "Odontosyllis fulgurans” | BT27 Species. P1 is identified by:  E42 Identifier |
| commonName | The nomenclatural code (or codes in the case of an ambiregnal name) under which the scientificName is constructed. Recommended best practice is to use a controlled vocabulary. | "Yellowfin Tuna"  “Τσιπούρα” | BT27 Species.  P1 is identified by:  E42 Identifier.  P2 has type:  “Common Name” |
| locality | The place that the common name is used. | "Greece” | E42 Identifier. is used in:  E53 Place |
| language | The language of the common name. | "Greek” | E42 Identifier.  P72 has language.  E56 language:E55 Type |

*Table 3*c*: Common Names Information*

## Micro-CT processes metadata

MicroCT (Micro Computed Tomography) is a technology that uses computer-processed X-rays to produce tomographic images (virtual 'slices') of specific areas of the scanned object, allowing the user to see inside without cutting. Digital geometry processing is used to generate a three-dimensional image of the inside of an object from a large series of two-dimensional radiographic images, taken around a single axis of rotation. It is a complex process that consists of information that is kept regarding the a series of events such as performspecimen’s preparation, the scanning process, the reconstruction and the postprocessing events. In the following table the metadata about these events are presented (Table 4a-4e).

**Metadata about Specimen preparation before scanning**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| specimenID | The ID of the specimen that will be prepared for scanning | "mCT-00001" | BC53 Specimen.  P1 is identified by: E41 Appellation |
| specimenLabel | Original label that assigned on specimen | "CalanusSt.307 Dorvilleidae CELB-5e-07" | BC53 Specimen. P1 is identified by: E41 Appellation |
| collectionID | The ID of the collection that the specimen belongs to | "urn:lsid:biocol.org:col:34871" | BC53 Specimen. P5 forms part of:  E78 Collection |
| specimenProvider | Person who provided the specimen that will be scanned | "SarahFaulwetter" "CharlotteWatson" | BC53 Specimen.  was provided by:  E39 Actor |
| providerInstitution | Institution which provided the specimen for scanning | "HCMR Darwin Museum of Natural History" | E39 Actor.  P107\_is\_current\_or\_former\_member\_of:  E40 Legal Body |
| specimenDescription | A verbatim description of the specimen, which allows to understand the nature of the specimen at a glance | "Copepod,whole animal"  "Polychaeta,anterior part" | BC53 Specimen. P3 has note: E62 String |
| species | The species that the specimen belongs to | "Odontosyllis fulgurans” | BC53 Specimen.  belongs to:  BT27 Species by |
| size\_mm | Maximum length of specimen that will be scanned | "3mm x 10mm" | BC53 Specimen. P43 has dimension:  E54 Dimension |
| preservationMedium | Current preservation medium of the specimen that will be scanned | "96%Ethanol"  "water"  "unknown" | E29 Design or Procedure |

*Table 4a: Metadata about preparation of specimen before scanning*

**Metadata about preparation during scanning**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **Marine TLO** |
| contrastEnhancementMethod | Short name of chemical that used during the preparation and scanning | "none"  "PTA"  "Iodine"  "HMDS" | E81 Transformation. P33 used specific technique:  E29 Design or Procedure |
| time-span | Date and Time of the duration of the preparation (e.g. the staining process) in the format YYYY-MM-DD and HH-MM | "9.30 am 10-10-2012 to 2.00 pm 10-10-2012" | Transformation. P4 has time-span: E52 Time-span |
| preparationNotes | Any other notes on the specimen preparation process | "2hours""24hours (overnight)""2x30 min""Object was inserted head-down into the pipette" | Transformation. P3 has note: E62 String |
| specimenID | A unique identifier for the specimen in the format: mCT-0000x (where x = incrementing number, always with preceding zeros) | "mCT-00001"  "mCT-00002" | Transformation. P124 transformed: BC53 Specimen.  P1 is identified by: E42 Identifier |

*Table 4b: Metadata about preparation during scanning*

**Metadata about scanning event**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| scanningEvent ID | The id of the scanning event. A unique code of the format scan-0000x (where x=incrementing number) | "scan-00001"  "scan-00002" | D2 Digitization Process. P48 has preferred identifier:E42 Identifier & D2 Digitization Process.  P2 has type. E55 Type: Scanning |
| scannedPart | (Body) part of the specimen that has been scanned | "whole specimen"  "whole bone"  "anterior part" | D2 Digitization Process. L1 digitized: E18 Physical Thing. P46 is composed of:  E18 Physical Thing |
| scannedBy | Person who performed the scanning | "SarahFaulwetter"  "ThanosDalianis" | D2 Digitization Process. P14 carried out by :  E39 Actor (BT9 Actor Type, BC8 Actor) |
| scanDate | Date of the begin of the scanning process in the format YYYY-MM-DD | "5/9/2012" | D2 Digitization Process. P4 has time-span: E52 Time Span *or L31 has starting date-time* |
| specimenID | A unique identifier for the specimen in the format: mCT-0000x (where x = incrementing number, always with preceding zeros) | "mCT-00001"  "mCT-00002" | D2 Digitization Process. L1F digitized:  BC53 Specimen.  P1 is identified by: E42 Identifier |
| producedDatasetID | The ID of the product of the scanning process | " mCT-0001.zip" | D2 Digitization Process. L20F has created: D9 Data Object |
| scanningDuration | The duration of the scan in the format HH:MM | "1:25:53 πμ" | D2 Digitization Process. L61 was ongoing at: Literal |
| sampleHolder | A description of the device that holds the sample | "yellow pipette tip"  "SkyScan plastic tube"  "styrofoam cylinder" | D2 Digitization Process. P16 used specific object: E70 Thing |
| digitalDevice | The Digital Device that was used for the scanning event | "08E01106" | D2 Digitization Process. L12 happened on device : D8 Digital Device (BC58 Digital Device) |
| digitalDeviceType | The type of the Digital Device that was used for the scanning event | "Skyscann" | D2 Digitization Process. L12 happened on device : D8 Digital Device. P2 has type:  E55 Type (BT13 Digital Object Type) |
| voltage\_kV | The voltage in kilovolt (kV)that used for scanning | "60" | D8 Digital Device.  P3 has note |
| current\_uA | The current in µAmpere that used for scanning | "200" | D8 Digital Device.  P3 has note |
| filter | Filter that used for scanning | "Aluminium" | D8 Digital Device.  P3 has note |
| zoom\_um | Resolution of the scan in µm (Zoom level) | "1,24" | D8 Digital Device.  P3 has note |
| cameraResolution | Camera resolution settings for scanning | "4000" | D8 Digital Device.  P3 has note |
| averaging | Frame averaging value that used for scanning | "3" | D8 Digital Device.  P3 has note |
| randomMovement | Random movement value that used for scanning | "0"  "10" | D8 Digital Device.  P3 has note |
| 360 degrees | 360° or 180° that used for scanning | "180" "360" | D8 Digital Device.  P3 has note |
| exposureTime\_ms | Exposure time in milliseconds that used for scanning | "125" | D8 Digital Device.  P3 has note |
| oversizesettings | Number of oversize parts (vertical & horizontal) that used for scanning | "2 vertical" | D8 Digital Device.  P3 has note |
| scanNotes | Any other notes on the scan | "specimen moved during scan" "scan too dark"  "projection images are missing" | D2 Digitization Process. P3 has note: E62 String |

*Table 4c: Metadata about scanning event*

**Metadata about Reconstruction Event**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| reconstructionEvent ID | The ID of the reconstruction event. A unique identifier in the form scan-0000x\_rec-0y | "scan-00001\_rec-01""scan-00002\_rec-01""scan-00040\_rec-02" | D3 Formal Derivation. P48 has preferred identifier:E42 Identifier & D3 Formal Derivation.  P2 has type. E55 Type: Reconstruction |
| reconScope | Reasonforreconstruction | "first recon""first 3 attempts in very bad quality, noise…""recon into different file format" | D3 Formal Derivation:  P17 was motivated by or P3 has note |
| reconBy | Person performing the reconstruction | "Sarah Faulwetter" | D3 Formal Derivation.  P14 carried out by :  E39 Actor (BT9 Actor Type, BC8 Actor) |
| inputDatasetID | The ID of the input of the reconstruction process | " mCT-0001.zip" | D3 Formal Derivation. L21F used derivation source:  D9 Data Object |
| producedDatasetID | The ID of the product of the reconstruction process | " mCT-0001-recon0001.zip" | D3 Formal Derivation. L22F created derivative: D9 Data Object |
| reconDate | Date of reconstruction in the format YYYY-MM-DD | "26/4/2012" | D3 Formal Derivation. P4 has time-span :  E52 Time-span |
| reconDuration | Duration of the reconstruction in the format HH:MM | "0:09:10" | D3 Formal Derivation.  P4 has time-span :  E52 Time-span |
| reconNotes | Any other notes and comments concerning the reconstruction | "Reconstruction files are absent"  "reconstruction in tiff format" | D3 Formal Derivation. P3 has note: E62 String |

*Table 4d: Metadata about Reconstruction Event*

**Metadata about Post Proccessing Event**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| postprocessingEventID | The ID of the postprocessing event | "Post Processing 001" | D3 Formal Derivation. P134 continued:  D3 Formal Derivation: P48 has preferred identifier:E42 Identifier & D3 Formal Derivation.  P2 has type. E55 Type: PostProcessing |
| subvolumeID | The product of the post processing event | "scan-00001\_rec-01\_voi-01"9"scan-00035\_rec-01\_voi-01""scan-00051\_rec-01\_voi-03" | D3 Formal Derivation. L22 created derivative:  D1 Digital Object. P1 is identified by: E42 Identifier |
| inputDatasetID | The ID of the input of the postprocessing process | " mCT-0001-recon0001.zip" | D3 Formal Derivation. L21F used derivation source:  D9 Data Object |
| subvolumeCreatedBy | The person who carried out the event | "SarahFaulwetter" | D3 Formal Derivation. P14 carried out by:  E39 Actor ( BT9 Actor Type, BC8 Actor) |
| subvolumeNotes | Any other notes and comments concerning the subvolume | "original recon misaligned, was reconstructed again"  "Subvolume used for creating surface models" | D3 Formal Derivation. P3 has note: E62 String |

*Table 4e: Metadata about PostProcessingEvent*

## Environmental Measurement Metadata

The following table illustrates metadata about the sample taking measurement event, such as the type of the dimension that was measured during the sample taking measurement event, the value of the dimension that was measured, etc..(Table 5)

**Metadata about Environmental Measurement Event**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| sample Taking Measurement Event ID | The ID of the sample taking measurement event | Sphaerosyllis levantina specimen  length measurement | S3 Measurement by Sampling. P48 has preferred identifier:E42 Identifier  &  S3 Measurement by Sampling.  P2 has type.  E55 Type: Environmental Measurement |
| measurementBy | The person that carried out the sample taking measurement event | "SarahFaulwetter" | S3 Measurement by Sampling.   P14 carried out by :  E39 Actor (BT9 Actor Type, BC8 Actor) |
| measurementDate | The date of the sample taking measurement event | "17/9/2012" | S3 Measurement by Sampling.  P4 has time-span :  E52 Time-span |
| locality | The place of the sample taking measurement event | "Alykes" | S3 Measurement by Sampling.  P7 took place at. E53 Place |
| observedDimension | The dimension that was measured during the sample taking measurement event | "totallength" | S3 Measurement by Sampling.  P40 observed dimension : E54 Dimension |
| dimensionType | The type of the dimension that was measured during the sample taking measurement event | "totallength" | S3 Measurement by Sampling.  P40 observed dimension : E54 Dimension. P2 has type: E55 Type |
| dimensionValue | The value of the dimension that was measured during the sample taking measurement event | "1.2" | S3 Measurement by Sampling.  P40 observed dimension : E54 Dimension. P90 has value:  E60 Number |
| dimensionUnit | The unit of the value of the dimension that was measured during the sample taking measurement event | "mm" | S3 Measurement by Sampling.  P40 observed dimension : E54 Dimension. P91 has unit :  E58 Measurement Unit |

*Table 5: Metadata about Sample Taking Measurement Event*

## Specimen Metadata

Specimen is an individual, item, or part representative of a class, genus, or whole.A biological specimen (also called a biospecimen) is a biological laboratory specimen held by a biorepository for research. Such a specimen would be taken by sampling so as to be representative of any other specimen taken from the source of the specimen. When biological specimens are stored, ideally they remain equivalent to freshly-collected specimens for the purposes of research. Here, the metadata are related to transformation event, the act or an instance of transforming an animal to specimen (Table 6).

**Metadata about Specimen Transformation Event**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| organismID | The ID of the individual from which the specimen was produced | urn:lsid:zoobank.org:act:9CEE8F90-9596-49F6-AA22-BB79C0E816D9 | E81 Transformation. P124 transformed:  BC38 Biotic Element. P48 has preferred identifier:E42 Identifier |
| transformationEvent ID | The id of the transformation event that transformed the individual into specimen | Sphaero-lev-00 | E81 Transformation.  P1 is identified by: E42 Identifier  &  E81 Transformation.  P2 has type.  E55 Type:  Specimen Creation |
| specimenID | The ID of the specimen that has transformed to | Sphaerosyllis-levantina-ALA-IL-7-Oct.2009 | E81 Transformation. P123 resulted in:  BC53 Specimen.  P1 is identified by: E42 Identifier |
| collectionID | The ID of the collection that the specimen belongs to | "urn:lsid:biocol.org:col:34871" | BC53 Specimen. P5 forms part of:  E78 Collection |
| methodID | The ID of method that used for transformation event | Preserved in ethanol | E81 Transformation. P33 used specific technique:  E29 Design or Procedure |
| actor | The person that carried out the transformation event | "Sarah Faulwetter" | E81 Transformation. P14 carried out by :  E21 Person (BT9 Actor Type, BC8 Actor) |
| time-span | The date of the transformation event | "17/9/2012" | E81 Transformation:  P4 has time-span: E52 Time-span |
| species | The species that the specimen belongs to | "Odontosyllis fulgurans" | BC53 Specimen. belongs to: BT27 Species |

*Table 6: Metadata about Transformation Event*

## Specimen Collection Metadata

Metadata about the creation event of specimen collections are presented below (Table 7). Especially, we define the actor that owns, keeps the specimen collection, the contributor and the curator of the specimen collection, etc..

**Metadata about Creation Event of Specimen Collections**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| collection ID | The ID of the collection that the specimen belongs to | "urn:lsid:biocol.org:col:34871" | BC42 Collection. P1 is identified by: E42 Identifier |
| collectionTitle | The title of the Collection that the specimen belongs to | "NAGISA" | BC42 Collection . P1 is identified by: BC19 Title |
| creationEvent ID | The ID of the creation event of the specimen collection | Sphaerosyllis-levantina:col-ALA-IL-7-Oct.2009-col | E56 Creation. P1 is identified by:E42 Identifier |
| actor | The person that created the specimen collection | "SarahFaulwetter" | E56 Creation.  P14 carried out by : E39 Actor(BT9 Actor Type, BC8 Actor) |
| eventDate | The date of the creation of the specimen collection | "17/9/2012" | E56 Creation. P4 has time-span : E52 Time-span |
| owner | The actor that owns the specimen collection | "HCMR" | BC42 Collection.  P52 has current owner :E39 Actor |
| keeper | The actor that keeps the specimen collection | "NHMC" | BC42 Collection. P52 has current keeper :E39 Actor |
| curator | The curator of the specimen collection that created | "SarahFaulwetter" | E56 Creation.  P14 is carried out by: E21 Person (BT9 Actor Type, BC8 Actor).  P14.1 in the role of:  E55 Type: Curator |
| contactPoint | The contact point of the curator of the data collection | [sarifa@hcmr.gr](mailto:sarifa@hcmr.gr) | E59 Actor. P76 has contact point: E51 Contact Point |
| remarks | A description of the specimen collection that created | Text | BC42 Collection.  P147 curated : E78 Collection.  P3 has note |

*Table 7: Metadata about Creation Event of Specimen Collections*

## Data Collection Metadata

The metadata that are kept for data collections are about the creation event of the data collection, the publication event, the embargo assignment and the human factor of the creation and publishing processes.

**Metadata about Data Collections**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| dataset ID | The id of the dataset. | "d50024ac-5268-477e-8559-933779553b34" | BC21 Data Set. P1 is identified by: E42 Identifier |
| creationEvent ID | The id of the creation event of the data collection | " d50024ac-5268-477e-8559-933779553b34\_creation" | E56 Creation. P1 is identified by: E42 Identifier |
| parentDataset ID | The ID of the collection of datasets that includes the described dataset | "Invertebrates" | BC21 Dataset. P5 forms part of:  BC21 Dataset |
| datasetTitle | The title of the dataset . | "EasternMedSyllids2" | BC21 Data Set.  P1 is identified by: E42 Identifier |
| creationDate | The date of the creation of the data collection | "17/9/2012" | E56 Creation. P4 has time-span:  E52 Time-span |
| creator | The person that created the data collection | "SarahFaulwetter" | E56 Creation.  P14 carried out by:  E39 Actor (BT9 Actor Type, BC8 Actor) |
| contributor | The contributor of the data collection | "SarahFaulwetter" "HCMR" | E56 Creation.  P14 is carried out by: E21 Person (BT9 Actor Type, BC8 Actor).  P14.1 in the role of:  E55 Type: Contributor |
| owner | The actor that owns the data collection | "HCMR" | BC42 Collection.  P52 has current owner: E39 Actor |
| publisher | The actor that published the data collection | "Sarah Faulwetter" | E56 Creation.  P14 is carried out by: E21 Person (BT9 Actor Type, BC8 Actor).  P14.1 in the role of:  E55 Type: Publisher |
| publicationDate | The date of the publication of the data collection | "17/9/2012" | E56 Creation. P4 has time-span:  E52 Time-span |
| keeper | The actor that keeps the data collection | "NHMC" | BC42 Collection.  P52 has current keeper: E39 Actor |
| curator | The curator of the data collection | "Sarah Faulwetter" | E56 Creation.  P14 is carried out by: E21 Person (BT9 Actor Type, BC8 Actor).  P14.1 in the role of:  E55 Type: Curator |
| contactPoint | The contact point of the curator of the data collection | [sarifa@hcmr.gr](mailto:sarifa@hcmr.gr) | E59 Actor. P76 has contact point: E51 Contact Point |
| accessRights | Information about who can access the rights of the data collection that created | http://creativecommons.org/licenses/by/3.0/ -- http://www.opendatacommons.org/licenses/by/1.0/ | BC42 Collection.  P104 is subject to : E30 Right |
| rightHolder | The holder of the rights of the data collection | "Israel Oceanographic &Limnological Research Ltd" | E30 Rights.  P105 right held by):  E39 Actor |
| access Method | The method that used to access the data collection | "download the data collection from www.collectionofhcmr.com" | BC 21 Dataset. P3 has note:  E62 String |
| description | A description of the data collection that created | "Text" | BC 21 Dataset. P3 has note:  E62 String |
| attributeAssigmentEvent ID | The id of the creation event that assigned an embargo state on the data collection | " d50024ac-5268-477e-8559-933779553b34 \_attribute\_assignment" | E13 Attribute Assignment. P1 is identified by: E42 Identifier |
| embargoState | This denotes if the datacollection is on or out of embargo. | “One Embargo” | E13 Attribute Assignment. P42\_assigned |
| embargoPeriod | The period that the data collection is on embargo. | “ 2010-2014 ” | E13 Attribute Assignment.  P4 has time-span:  E52 Time-span |
| image | The logo of the data collection | “ThunnusCollection.jpg” | BC 21 Dataset. P62 is depictedby |

*Table 8: Metadata about Creation Event of Data Collections*

## Genetic Events Metadata

Genetics is the branch of biology that deals with heredity, especially the mechanisms of heredity transmission and the variation of inherited characteristics among similar or related organisms. Here, metadata about sampling and DNA sequence extraction events are presented (Table 9a-9b). DNA sequencing, one of the most fundamental technologies developed to study genetics, allows researchers to determine the sequence of nucleotides in DNA fragments.

**Metadata about Sampling Event**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| sampleTakingEvent ID | The ID of the sample taking event | "MaterialExtraction 002" | S2 Sample Taking. P48 has preferred identifier:E42 Identifier & S2 Sample Taking. P2 has type. E55 Type:  Genetics Sampling |
| collection\_date | The date of sampling. Use ISO8601 compliant format | "2011-02-18" | S2 Sample Taking.  P4 has time-span: E52 Time-span |
| collection\_time | The local time of sampling. Use ISO8601 compliantformat | "11:00" | S2 Sample Taking.  P4 has time-span:  E52 Time-span |
| country | The geographical origin of the sample as defined by the country or sea name. | "Greece" | S2 Sample Taking. P7 took place at: E53 Place |
| locality | The geographical origin of the sample as defined by the specific local region name | "Logarou lagoon (Amvrakikos Gulf)"2"Rodia lagoon (Amvrakikos Gulf)" | S2 Sample Taking.  P7 took place at:. E53 Place. P1 is identified by: E41 Appellation |
| sample ID | The ID of the sample that was extracted by specimen | "Station in the channel connecting the Rodia lagoon to the Amvrakikos Gulf (replicate A)" "Station inside the Rodia lagoon (replicate B)" "Individual material" | S2 Sample Taking.  O5 removed:  S13 Sample. P1 is identified by: E42 Identifier |
| samp\_collection\_device | The device or method that used during sample collection | "Sequencer 08E01106" | S2 Sample Taking. P16 used specific object : D8 Digital Device |
| samp\_size | Amount or size of sample that collected | "Amount = 5"  "Size = 16KB" | S2 Sample Taking. O5 removed:  S13 Sample. P43 has dimension :E54 Dimension |

*Table 9a: Metadata about Sampling Event*

**Metadata about DNA Sequencing Event**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| DNA SequencingProcess ID | The ID of the sequencing event | "tatutggugaatt-1753" | D2 Digitization Process. P48 has preferred identifier:E42 Identifier  & D2 Digitization Process. P2 has type. E55 Type:  DNA Sequencing |
| project\_name | Name of the project within which the sequencing was organized | "BiodiversityPatternsinAmvrakikos" | BC21 Dataset. P1 is identified by: E42 Identifier |
| institution | Organizationcontact about DNA sequence extraction event | "Hellenic Centre for Marine Research" | BC13 Organisation |
| project\_description | Description of the project within which the DNAsequencing was organized | "Chain-termination method" "Shotgun sequencing method" | BC21 Dataset. P3 has note: E62 String |
| SequencingProcessProduct ID | The ID of the product of the sequencing process | " 6598-atuggatuatgaut-8975" | D2 Digitization Process. L22F created derivative: D9 Data Object |

*Table 9b: Metadata about Sequencing Event*

## Measurement Event Metadata

Measurement event is the event of measuring dimensions of specimens or individuals that belong to a species such as length, weight, etc.

**Metadata about Measurement event**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| measurementEvent ID | The ID of the measurement event |  | E16 Measurement. P48 has preferred identifier:E42 Identifier & E16 Measurement. P2 has type: E55 Type:  Species Measurement |
| measuredBy | The person that carried out the measurement event | "SarahFaulwetter" | E16 Measurement . P14 carried out by: E39 Actor |
| measurementDate | The date of the measurement event | "17/9/2012" | E16 Measurement. P4 has time-span:  E52 Time-span |
| specimenID | The id of the specimen that was measured | Sphaerosyllis-levantina-ALA-IL-7-Oct.2009 | E16 Measurement. P39 measured:  BC53 Specimen. P1 is identified by: E42 Identifier |
| species | The species that the specimen that was measured belongs to, or the species that the attribute was assigned at | "Odontosyllisfulgurans" | BC53 Specimen. belongs to: BT27 Species |
| observedDimension | The dimension that was measured | "totallength" | E16 Measurement:  P40 observed dimension :  E54 Dimension |
| DimensionType | The type of the dimension that was measured | "totallength" | E54 Dimension:P2 has type:E55 Type |
| DimensionValue | The value of the dimension that was measured | "1.2" | E54 Dimension: P90 has value:E60 Number |
| DimensionUnit | The unit of the value of the dimension that was measured | "mm" | E54 Dimension: P91 has unit : E58 Measurement Unit |
|  |  |  |  |

*Table 10: Metadata about Measurement event*

## Morphological Characteristics Metadata

The evolutionary development of form in an organism or part of an organism refers to the quantitative analysis of form, a concept that encompasses size and shape. Morpholofical analyses are commonly performed on organisms, and are useful in analyzing their fossil record, the impact of mutations on shape, developmental changes in form, covariance between ecological factors and shape, as well for estimating quantitative-genetic parameters of shape.

**Metadata about Morphological Characteristics Assignment Event**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| attributeAssignmentEventURI | The ID of the measurement event |  | E13 Attribute Assignment. P48 has preferred identifier:E42 Identifier & E13 Attribute Assignment. P2 has type: E55 Type:  Morphometrics Assigment |
| actor | The person that carried out the measurement event | "Sarah Faulwetter" | E13 Attribute Assignment.  P14 carried out by:  E39 Actor |
| time-span | The date of the measurement event | "17/9/2012" | E13 Attribute Assignment. P4 has time-span:  E52 Time-span |
| species | The species that the characteristic was attributed to. | "Odontosyllisfulgurans" | E13 Attribute Assignment. P140 assigned attribute to: BT27 Species |
| observedDimension | The dimension that was measured | "color" | E13 Attribute Assignment. P37 assigned:  E54 Dimension |
| DimensionType | The type of the dimension that was measured | "maxlength" | E54 Dimension:  P2 has type:  E55 Type |
| DimensionValue | The value of the dimension that was measured | "1.2" | E54 Dimension:  P90 has value: E60 Number |
| DimensionUnit | The unit of the value of the dimension that was measured | "mm" | E54 Dimension:  P91 has unit: E58 Measurement Unit |

*Table 11: Metadata about Morphological Characteristics Assignment event*

## Statistical Metadata

Statistics is the study of the collection, [analysis](http://en.wikipedia.org/wiki/Analysis), interpretation, presentation, and organization of [data](http://en.wikipedia.org/wiki/Data). Here, we present metadata about statistical analysis of data, such as the person who carried out this statistical analysis, the methods and the protocols which have been used to the statistical analysis of data, etc..(Table 12).

**Metadata about Statistical Analysis of Data**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata** | **Definition** | **Example(s)** | **CRM family models** |
| statistical Analysis Event ID | The ID of the statistical analysis event | “statistical-analysis-001” | S6 Data Evaluation. P1 is identified by: E42 Identifier & S6 Data Evaluation. P2 has type: E55 Type:  Statistical Analysis |
| determinedBy | The person who carried out the statistical analysis of data | "SarahFaulwetter" | S6 Data Evaluation.  P14 carried out by :  E39 Actor |
| time-Span | The date of statistical analysis of data | "17/9/2012" | S6 Data Evaluation.P4 has time-span :  E52 Time-span |
| statisticalRemarks | Any note relative to the statistical analysis of data | "smoothing"  "testingandvisualizing | S6 Data Evaluation.  P3 has note: E62 String |
| statisticalMethod | The methods and the protocols which have been used to the statistical analysis of data | "Analysis of Variance" "Mann–Whitney U test" | S6 Data Evaluation.  P32 used general technique :  E55 Type |
| observedDimension | The dimension that was measured | "Average Specimen Length" | S6 Data Evaluation. O10 assigned dimension:  E54 Dimension |
| DimensionType | The type of the dimension that was measured | " Average Specimen Length " | E54 Dimension:  P2 has type:  E55 Type |
| DimensionValue | The value of the dimension that was measured | "5" | E54 Dimension:  P90 has value: E60 Number |
| DimensionUnit | The unit of the value of the dimension that was measured | "mm" | E54 Dimension:  P91 has unit: E58 Measurement Unit |

*Table 12: Metadata about Statistical Analysis of Data*

# Deliverable 2.2 Report

*Starting Date:* 1/1/2013

*Ending Date:* 30/9/2015

*Deliverable First Version Date:* 31/12/2014

*FORTH Participants:* Athina Kritsotaki: 1/1/2014 – now

Nikos Minadakis: 1/4/2013 – now

Vasiliki Tsimbida: 1/10/2014 - now

*Final Version Progress:* 60%

**1/1/2013 – 31/12/2014 Progress Report:**

During months 1-12 biodiversity datasets of the partners were collected and analyzed. The main metadata categories were recognized and lists of the competency queries were created in collaboration with the domain experts for each category. A number of meetings between FORTH, HCMR and domain experts from other institutions took place in order to come to an agreement on the kept metadata and the format of them. Templates for each metadata category were created in order to assist the metadata submission progress.

During months 13-24 the analysis and collection of biodiversity datasets and requirements continued and furthermore the mapping between the source’s metadata and the semantic models of the LW Greece Infrastructure begun. A number of meetings and presentations regarding these mappings took place mainly between FORTH and HCMR. The semantic models were extended according to the needs of the related competency queries and the form of the source’s metadata.

**Next Steps until 30/9/2015:**

Until the end version of the deliverable the remaining datasets will be analyzed and the process of collecting metadata, map them to the semantic model and extend the model accordingly will continue. The metadata categories and fields will come to the final format such as the related templates. Furthermore any change that has to do with the needs of the deliverable 2.4 of data services will take place in order to keep an alignment between the different components of the infrastructure.

1. Classes prefixes E and properties prefixes P refer to cidoc-crm  
    Classes prefixes S and properties prefixes O refer to crm-sci

   Classes prefixes D and properties prefixes L refer to crm-dig

   Classes prefixes BC/BT and properties prefixes LC/LX/LTrefer to marineTLO [↑](#footnote-ref-1)