Publishing sample data using the GBIF IPT

# Expressing sample data in Darwin Core

The Darwin Core vocabulary already provides a rich set of terms, organised into several classes (e.g., Occurrence, Event, Location, Taxon, Identification, MeasurementOrFact). Many of these terms are relevant for describing sample based data. Synthesising several sources of inputs (GBIF organised workshop on sample data, May 2013; discussions on the TDWG mailing list; discussions on the EU BON mailing list), a small set of descriptors relating to the sample data were identified as essential, most of which are already present in the DwC vocabulary. These terms are:

1. **eventID**: an identifier for the set of information associated with an Event; may be a global unique identifier or an identifier specific to the data set.
2. **eventSeriesID**: an identifier for a set of events that are associated in some way, e.g., a monitoring series; may be a global unique identifier or an identifier specific to the series.
3. **samplingProtocol**:the name of, reference to, or description of the method or protocol used during an Event.
4. **samplingEffort**: a numeric value for the time duration, length, area or volume involved in the sampling.
5. **samplingUnit**: the unit of measurement used for reporting the quantity in the sample.
6. **samplingGeometry**: an indication of what kind of space was sampled; select from point, line, area or volume.
7. **quantity**: the number of the quantityType (e.g., individuals, biomass, biovolume) per methodUnit or a percentage measure recorded for the sample.an integer number representing the number of individuals or a percentage measure recorded for the sample.
8. **quantityType**: the entity being referred to by quantity, e.g., individuals, biomass, %species.
9. **quantityUnit**: the unit of measurement being referred to by quantity.
10. **individualCount**: the number of individuals represented present at the time of the event.

Six of the ten terms which are closely inter-related through the sampling methodology are listed in Table 1 together with some example values.

Table . Terms relating to sampling methodology.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **samplingGeometry** | **samplingUnit** | **samplingEffort** | **quantity** | **quantityType** | **quantityUnit** | **individualCount** |
| Point  e.g. sensors, camera traps, pitfall traps | e.g., min | numeric time duration | number | * Individuals * Biomass * % Biomass * Biovolume * % Biovolume * % species * % coverage | e.g., gm  e.g., kg | integer |
| line | e.g., metre | numeric length |
| area | e.g., metre^2 | numeric surface |
| volume | e.g., metre^3 | numeric volume |

The value for quantityType (i.e., the entity being measured) is expected to be drawn from a small controlled vocabulary. Typical values include: Individuals, Biomass , % Biomass, Biovolume, % Biovolume, % species, % coverage.

The samplingGeometry determines the samplingUnit. Points are associated with temporal units (e.g., minute, hour, day), line with length units (e.g., metre, kilometre), area with areal units (e.g., metre^2, kilometre^2, hectare) and volume with volumetric units (e.g., litre, metre^3).

Of the complete set of ten terms identified, four already exist in DwC (eventID, samplingProtocol, samplingEffort, individualCount). Three other terms (quantity, quantityType, quantityUnit) can be mapped (Table 2) to equivalent MeasurementOrFact terms (measurementValue, measurementType, measurementUnit), leaving just three (samplingGeometry, samplingUnit, eventSeriesID) to be created.

Table 2. Mapping MeasurementOrFact terms.

|  |  |
| --- | --- |
| **measurementValue**: the value of the measurement, fact, characteristic, or assertion. | **quantity**: an integer number representing the number of individuals or a percentage measure recorded for the sample. |
| **measurementType**: the nature of the measurement, fact, characteristic, or assertion. Recommended best practice is to use a controlled vocabulary. | **quantityType**: the type of measurement being referred to by quantity. |
| **measurementUnit**: the units associated with the measurementValue. Recommended best practice is to use the International System of Units (SI). | **quantityUnit**: the unit of measurement used for reporting the quantity in the sample |

# Darwin Core Archive

|  |  |
| --- | --- |
| **Event Core** | **Occurrence Extension** |
| eventID | eventID |
| eventSeriesID\* | measurementValue |
| samplingProtocol | measurementType |
| samplingEffort | measurementUnit |
| samplingUnit\* | individualCount |
| samplingGeometry\* |  |

The \* symbol indicates proposed new terms.

# Examples

## Rhine Main Observatory Aquatic Invertebrates Biodiversity

**Extension (event-occurrence) table**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| EventID | Scientific Name | Measurement  Value | Measurement  Unit | Measurement  Type | Individual  Count | … |
| C\_1428 | Baetis rhodani | 14 |  | individuals | 70 |  |
| C\_1428 | Ephemera danica | 12.2 |  | individuals | 61 |  |
| C\_1428 | Gyraulus albus | 3.4 |  | individuals | 17 |  |

**Core (Event) table**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| EventID | Event  SeriesID | Sampling  Protocol | Sampling  Effort | Sampling  Geometry | Sampling Unit | Event  Date | Location | Lat | Long |
| C\_1428 | AN-358 | AQEM | 5 | area | Metre  (Or  Metre^2) | 21/06/2006 | Kinzig O3 Rothenbergen | 50.18689 | 9.100369 |

**Ephemera danica** : 12.2 individuals per sq metre. In total, 5 sq metres were sampled (*samplingeffort*; *samplingUnit*) and the total count for the species was 61 (*individualCount*). As the *measurementType* is “individuals”, there is no *measurementUnit* in this case.

## Shore transect survey

**Extension (event-occurrence) table**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| EventID | Scientific Name | Measurement  Value | Measurement  Unit | Measurement  Type | Individual  Count | … |
| B\_2488 | Chondrus crispus | 5 | grams | biomass |  |  |
| B\_2488 | Nucella lapillus | 12.8 |  | %species | 25 |  |
| B\_2488 | Carcinus maenas | 3.1 |  | %species | 7 |  |
| B\_2488 | Mastocarpus stellatus | 31 |  | %coverage |  |  |

**Core (Event) table**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| EventID | Event  SeriesID | Sampling  Protocol | Sampling  Effort | Sampling  Geometry | Sampling Unit | Event  Date | Location | Lat | Long |
| B\_2488 |  | PISCO | 3 | area | metre | 08/05/2001 | Lough Hyne | 51.50 | -9.30 |

**Chondrus crispus**: 5 grams (biomass) per sq metre. In total, 3 sq metres were sampled (*samplingEffort*; *samplingUnit*)

**Nucella lapillus**: 12.8 (%species). In total, 3 sq metres were sampled (*samplingEffort*; *samplingUnit*) and the *individualCount* for the species was 25. %species is calculated from *individualCount* as a % of the summed individualCounts of all the species present.

**Mastocarpus stellatus**: 31 (%coverage). In total, 3 sq metres were sampled (*samplingEffort*; *samplingUnit*).