Slides available at: https://bit.ly/gbifEu2018_new_datatypes

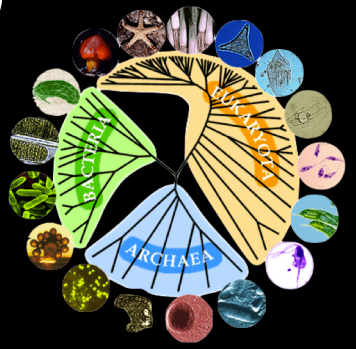
Event core and new data types in GBIF

European GBIF Nodes Meeting in Tallinn 14-16 May 2018

Slides, CC-BY Dag Endresen, GBIF.no

New data types

- Multimedia (Audubon Core)
- Collections (NCD)
- Taxon concepts, taxon traits (Plinian Core)
- Molecular genotype data (GGBN)
- Experiment or fact, organism traits
- …
- New data types from Darwin Core not yet supported in GBIF?



Darwin Core : identifier terms

Record-level Terms | **dcterms:type** | dcterms:modified | dcterms:language | dcterms:license | dcterms:rightsHolder | dcterms:accessRights | dcterms:bibliographicCitation | dcterms:references | **institutionID** | **collectionID** | **datasetID** | **institutionCode** | **collectionCode** | **datasetName** | ownerInstitutionCode | **basisOfRecord** | informationWithheld | dataGeneralizations | dynamicProperties

Occurrence | **occurrenceID** | **catalogNumber** | **recordNumber** | recordedBy | individualCount | organismQuantity | organismQuantityType | sex | lifeStage | reproductiveCondition | behavior | establishmentMeans | occurrenceStatus | preparations | disposition | associatedMedia | associatedReferences | associatedSequences | associatedTaxa | otherCatalogNumbers | occurrenceRemarks

Organism | **organismID** | **organismName** | organismScope | associatedOrganisms | previousIdentifications | organismRemarks

MaterialSample | LivingSpecimen | PreservedSpecimen | FossilSpecimen | **materialSampleID**

Event | HumanObservation | MachineObservation | **eventID** | **parentEventID** | **fieldNumber** | eventDate | eventTime | startDayOfYear | endDayOfYear | year | month | day | verbatimEventDate | habitat | samplingProtocol | sampleSizeValue | sampleSizeUnit | samplingEffort | fieldNotes | eventRemarks

Location | **locationID** | **higherGeographyID** | **higherGeography** | continent | waterBody | islandGroup | island | country | countryCode | stateProvince | county | municipality | **locality** | verbatimLocality | minimumElevationInMeters | maximumElevationInMeters | verbatimElevation | minimumDepthInMeters | maximumDepthInMeters | verbatimDepth | minimumDistanceAboveSurfaceInMeters | maximumDistanceAboveSurfaceInMeters | locationAccordingTo | locationRemarks | decimalLatitude | decimalLongitude | geodeticDatum | coordinateUncertaintyInMeters | coordinatePrecision | pointRadiusSpatialFit | verbatimCoordinates | verbatimLatitude | verbatimLongitude | verbatimCoordinateSystem | verbatimSRS | footprintWKT | footprintSRS | footprintSpatialFit | georeferencedBy | georeferencedDate | georeferenceProtocol | georeferenceSources | georeferenceVerificationStatus | georeferenceRemarks

GeologicalContext | **geologicalContextID** | earliestEonOrLowestEonothem | latestEonOrHighestEonothem | earliestEraOrLowestErathem | latestEraOrHighestErathem | earliestPeriodOrLowestSystem | latestPeriodOrHighestSystem | earliestEpochOrLowestSeries | latestEpochOrHighestSeries | earliestAgeOrLowestStage | latestAgeOrHighestStage | lowestBiostratigraphicZone | highestBiostratigraphicZone | lithostratigraphicTerms | group | formation | member | bed

Identification | **identificationID** | identificationQualifier | typeStatus | identifiedBy | dateIdentified | identificationReferences | identificationVerificationStatus | identificationRemarks

Taxon | **taxonID** | **scientificNameID** | **acceptedNameUsageID** | **parentNameUsageID** | **originalNameUsageID** | **nameAccordingToID** | **namePublishedInID** | **taxonConceptID** | **scientificName** | **acceptedNameUsage** | **parentNameUsage** | **originalNameUsage** | **nameAccordingTo** | **namePublishedIn** | namePublishedInYear | higherClassification | kingdom | phylum | class | order | family | genus | subgenus | specificEpithet | infraspecificEpithet | taxonRank | verbatimTaxonRank | scientificNameAuthorship | vernacularName | nomenclaturalCode | taxonomicStatus | nomenclaturalStatus | taxonRemarks

MeasurementOrFact | **measurementID** | measurementType | **measurementValue** | measurementAccuracy | measurementUnit | measurementDeterminedBy | measurementDeterminedDate | measurementMethod | measurementRemarks

ResourceRelationship | **resourceRelationshipID** | **resourceID** | **relatedResourceID** | **relationshipOfResource** | relationshipAccordingTo | relationshipEstablishedDate | relationshipRemarks

Darwin Core: identifier terms

Record-level Terms | dcterms:type | basisOfRecord | institutionID | collectionID | datasetID | institutionCode | collectionCode | datasetName

Occurrence | occurrenceID | catalogNumber | recordNumber

Organism | organismID | organismName

MaterialSample | LivingSpecomen | PreservedSpecimen | FossilSpecimen | materialSampleID

Event | HumanObservation | MachineObservation | eventID | parentEventID | fieldNumber

Location | locationID | higherGeographyID | higherGeography | locality | decimalLatitude | decimalLongitude

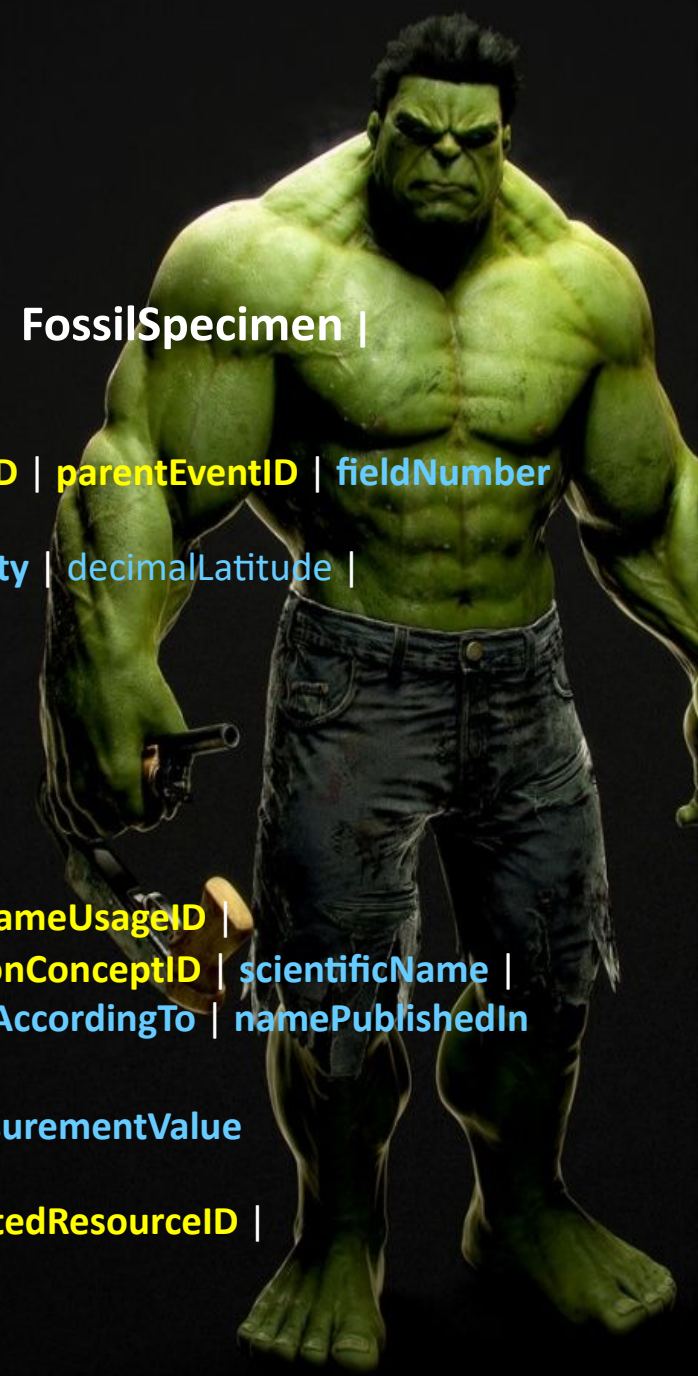
GeologicalContext | geologicalContextID | …

Identification | identificationID | …

Taxon | taxonID | scientificNameID | acceptedNameUsageID | parentNameUsageID | originalNameUsageID | nameAccordingToID | namePublishedInID | taxonConceptID | scientificName | acceptedNameUsage | parentNameUsage | originalNameUsage | nameAccordingTo | namePublishedIn

MeasurementOrFact | measurementID | measurementType | measurementValue

ResourceRelationship | resourceRelationshipID | resourceID | relatedResourceID | relationshipOfResource



Simple Darwin Core

«The Simple Darwin Core is simple in that it assumes (and allows) no structure beyond the concept of rows and columns, which might be thought of as attributes and their values, or fields and records.»

In other words: **de-normalized**.

Darwin Core RDF Guide (2015)

Biodiversity Information Standards TDWG

Introduction
References
Quick Reference Guide
Simple Darwin Core

RDF Guide
Table of Contents
Introduction
Implementation Guide
Term reference
Text Guide
XML Guide

Namespace Policy
Complete History
Decision History

Mapping to ABCD
Mapping to Old Versions

Darwin Core RDF Guide

Title: Darwin Core RDF Guide

Date Issued: 2015-03-27

Date Modified: 2015-06-15

Abstract: This guide is intended to facilitate the use of Darwin Core terms in the Resource Description Framework (RDF). It explains basic features of RDF and provides details of how to expose data in the form of RDF using Darwin Core terms and terms from other key vocabularies. It defines terms in the namespace http://rs.tdwg.org/dwc/iri/ which are intended for use excusively with non-literal objects.

Contributors: Steve Baskauf (TDWG RDF/OWL Task Group), John Wieczorek (TDWG Darwin Core Task Group), John Deck (Genomic Biodiversity Working Group), Campbell Webb (TDWG RDF/OWL Task Group), Paul J. Morris (Harvard University Herbaria/Museum of Comparative Zoölogy), Mark Schildhauer (National Center for Ecological Analysis and Synthesis) (KUNHM)

Legal: This document is governed by the standard legal, copyright, licensing provisions and disclaimers issued by the Taxonomic Databases Working Group.

Part of TDWG Standard: http://www.tdwg.org/standards/450/

Creator: Darwin Core and RDF/OWL Task Groups

Identifier: http://rs.tdwg.org/dwc/2015-06-02/terms/guides/rdf/

Latest Version: http://rs.tdwg.org/dwc/terms/guides/rdf/

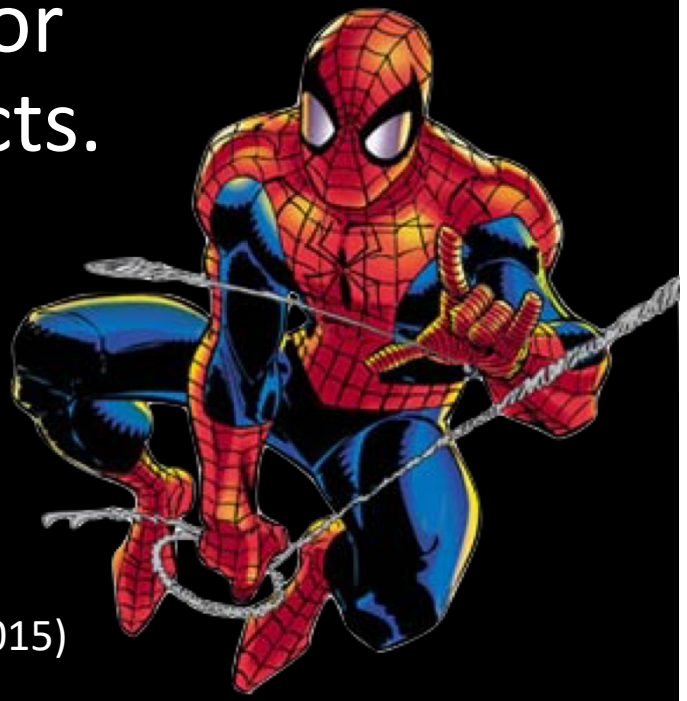
Replaces: http://rs.tdwg.org/dwc/2015-03-27/terms/guides/rdf/

Document Status: Current Standard

dwciri: namespace

The dwciri: namespace is intended for use exclusively with non-literal objects.

http://rs.tdwg.org/dwc/iri/



Baskauf S, Wieczorek J, Deck J, Webb C, Morris PJ, and Schildhauer M (2015) Darwin Core RDF Guide, Biodiversity Information Standards http://rs.tdwg.org/dwc/terms/guides/rdf/

Baskauf SJ, Wieczorek J, Deck J, and Webb CO (2016) Lessons learned from adapting the Darwin Core vocabulary standard for use in RDF. Semantic Web 7(6): 617-627, DOI: 10.3233/SW-150199

dwciri: and dwc: terms

- *"If a term in the **dwciri**: namespace has an analogue in the **dwc**: namespace having the same local name, the dwciri: term will have the same meaning as its dwc: counterpart, but is intended to have an IRI-reference object."*

RDF Guide & dwciri terms

New dwciri terms without corresponding terms in the dwc namespace

dwciri:inCollection | dwciri:toTaxon | dwciri:inDescribedPlace | dwciri:earliestGeochronologicalEra | dwciri:latestGeochronologicalEra | dwciri:fromLithostratigraphicUnit | dwciri:inDataset

dwciri terms with analogues in the dwc namespace

recordedBy | identifiedBy | georeferencedBy | measurementDeterminedBy | locationAccordingTo | | georeferenceProtocol | georeferenceSources | samplingProtocol | sex | lifeStage | reproductiveCondition | establishmentMeans | behavior | occurrenceStatus | disposition | verbatimCoordinateSystem | verbatimSRS | geodeticDatum | georeferenceVerificationStatus | footprintWKT | footprintSRS | lowestBiostratigraphicZone | highestBiostratigraphicZone | identificationVerificationStatus | identificationQualifier | preparations | typeStatus | measurementType | measurementUnit | measurementMethod | informationWithheld | dataGeneralizations | habitat | fieldNumber | fieldNotes | recordNumber

«IRI» terms imported from Dublin Core

dcterms:language | dcterms:license | dcterms:type |

dcterms:rightsHolder | dcterms:accessRights | dcterms:references

Required terms imported from outside of Darwin Core

rdf:type | dcterms:identifier | dcterms:relation

Darwin Core RDF Guide

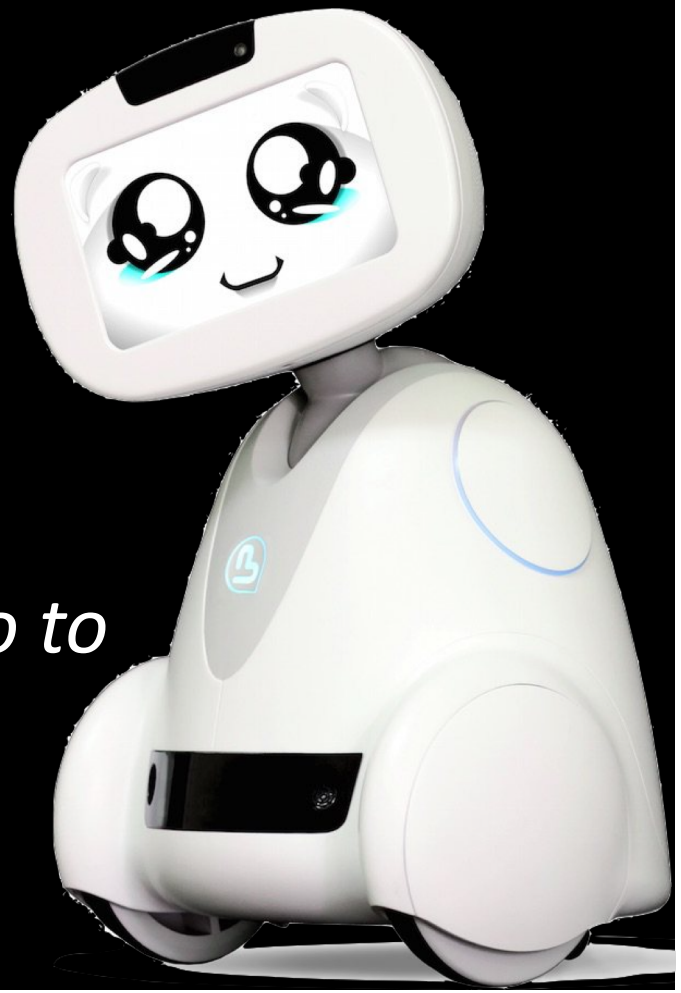
*«A number of Darwin Core properties specify that their values should be an **identifier**. There is significant ambiguity in the use of these properties because depending on the situation the value may be an **identifier for the subject resource itself** or it may be the **identifier for a resource that is related** in some way to the subject resource.»*

*«Darwin Core contains a number of "ID" terms intended to designate identifiers, e.g., dwc:**occurrenceID**, dwc:**identificationID**, dwc:**locationID**, etc. The "ID" terms provide two functions, specifying the **class** of the resource and indicating that value of the term is an **identifier**.»*

http://rs.tdwg.org/dwc/terms/guides/rdf/index.htm

Unintended consequences of **ID terms**

- In its normative definition, each Darwin Core ID term is declared to be rdfs:subPropertyOf dcterms:identifier.
- In RDF, the purpose of a subproperty declaration is to allow a client with reasoning capability to infer a triple containing a broader (and presumably more well-known) property.
- (*Applicable to RDF-serialization, less so to text-or XML-serialization*).



DwC RDF Guide ID term example 23

Example 23 : Turtle (incorrect usage)

```
<http://guid.mvz.org/identifications/23459>
  a dwc:Identification; dwc:identifiedBy "Richard Sage";
  dwc:dateIdentified "2000"^^xsd:gYear;
  dwc:taxonID
"http://lsid.tdwg.org/urn:lsid:catalogueoflife.org:taxon:d79c11aa-
29c1-102b-9a4a-00304854f820:col20120721".
```

possible incorrect inference:

```
  dcterms:identifier
"http://lsid.tdwg.org/urn:lsid:catalogueoflife.org:taxon:d79c11aa-
29c1-102b-9a4a-00304854f820:col20120721".
```

Darwin Core archive and Event dataset

Event core and measurements

- Event core (since 2015) opens up new functionality for data published in GBIF!
- Limitations of the star-schema is a frequently mentioned obstacle – of event core…
  - E.g. when connecting measurement or facts (MoF) only core entities can be described.

DwC archive – application schema

DwC-A Event (since 2015)

- http://rs.gbif.org/core/dwc_event_2016_06_21.xml

DwC-A Occurrence

- http://rs.gbif.org/core/dwc_occurrence_2015-07-02.xml

DwC-A Measurement or fact (want link to event and occ!)

- http://rs.gbif.org/extension/dwc/measurements_or_facts.xml

OBIS Extended measurement or fact (proposed 2017)

- http://rs.gbif.org/extension/obis/extended_measurement_or_fact.xml

DwC-A Resource relationship (possible solution?)

- http://rs.gbif.org/extension/dwc/resource_relation.xml

Use case

Use case : Measurements for both events and occurrences in the same dataset

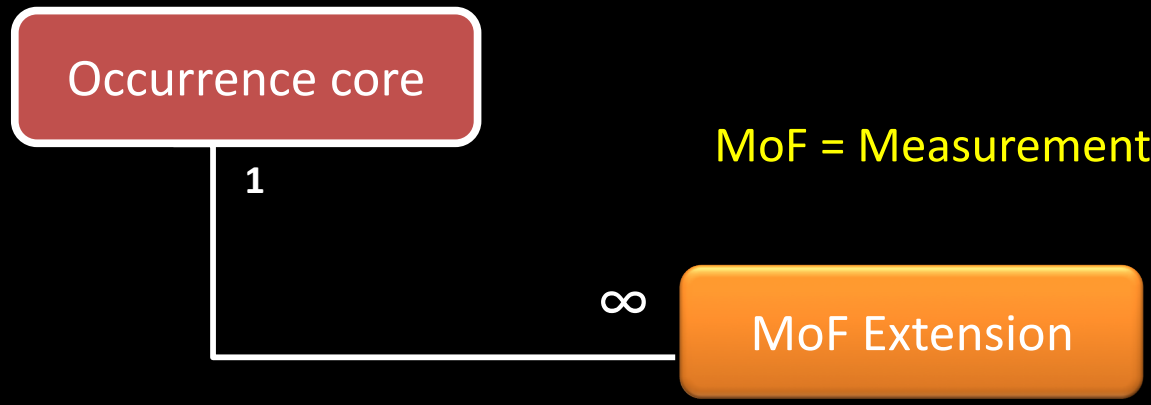
OBIS expeditions are sampling more data than only organism occurrences (e.g. environment).

--> development of a new extended measurement and fact extension (eMoF for DwC-archives)

Integrated Publishing Toolkit

Star Schema (DwC-archive)

International Oceanographic Data and Information Exchange



MoF = Measurement or Fact

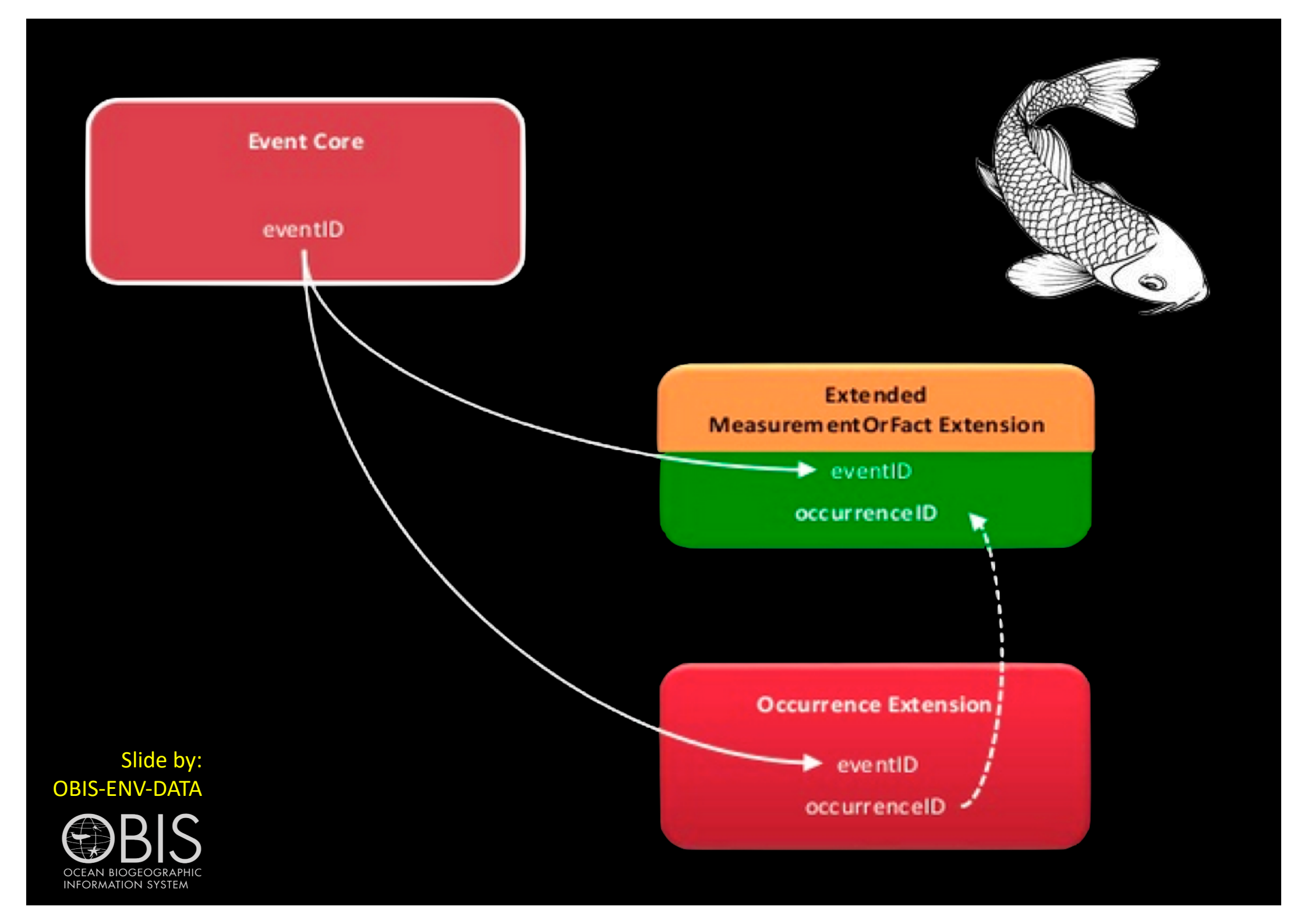
Here: MoF describes only the Occurrences

Sample event is the DwC core

Here: MoF describes only the Event

Slide by:
OBIS-ENV-DATA

OBIS
OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM


Event Core
eventID
Extended
MeasurementOrFact Extension
eventID
occurrenceID
Occurrence Extension
eventID
occurrenceID
Slide by:
OBIS-ENV-DATA
OBIS
OCEAN BIOGEOGRAPHIC
INFORMATION SYSTEM

eMoF propose minting three new terms

measurementType: The nature of the measurement or fact (MoF)

measurementTypeID (new): identifier for the measurementType (global unique identifier, URI)

http://rs.iobis.org/obis/terms/measurementTypeID

measurementValue: the value of the MoF

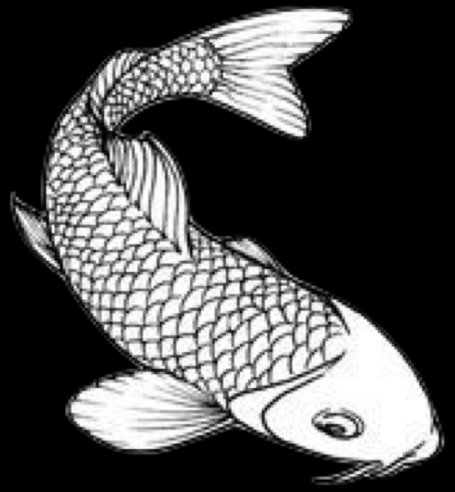
measurementValueID (new): identifier for facts reported as measurementValue (URI)

http://rs.iobis.org/obis/terms/measurementValueID

measurementUnit: measurement unit (SI unit; meter, gram, etc)

measurementUnitID (new): identifier for the measurementUnit (URI)

http://rs.iobis.org/obis/terms/measurementUnitID



Let's explore the Darwin Core RDF Guide first…

Darwin Core RDF Guide (2015)

Darwin Core RDF guide and the dwciri: namespace term already have alternatives for some of the new terms minted by the OBIS eMoF.

emof = http://rs.iobis.org/obis/terms/
dwciri = http://rs.tdwg.org/dwc/iri/

emof:measurementTypeID = dwciri:measurementType
emof:measurementUnitID = dwciri:measurementUnit
emof:measurementValueID (no analogue in dwciri - useful addition)

see also:
dwciri:measurementDeterminedBy
dwciri:measurementMethod
dwciri:samplingProtocol

DwC RDF Guide: http://rs.tdwg.org/dwc/terms/guides/rdf/index.htm#2.9_MeasurementOrFact_instances

Linking MoF instances to resources

- "*Darwin Core does not provide the object properties that would be required to describe precisely how the dwc:MeasurementOrFact instance was related to the dwc:Occurrence instance, the dwc:Organism whose occurrence was recorded, or the dwc:PreservedSpecimen that was collected*".
- Examples of the DwC RDF Guide use **dcterms:relation** to link MoF instances to the respective resources.

Baskauf *et al*. 2015, Darwin Core RDF Guide, chapter 2.9

DwC RDF Guide MoF example 31

Example 31 - Turtle

```
<http://mczbase.mcz.harvard.edu/guid/MCZ:Mamm:67231>
  a dwc:Occurrence;
  dcterms:relation [a dwc:MeasurementOrFact ;
    dwc:measurementType "tail length" ;
    dwciri:measurementType <http://purl.obolibrary.org/obo/VT_0002758> ;
    dwc:measurementValue "25"^^xsd:int ; dwc:measurementUnit "mm" ;
    dwciri:measurementUnit
         <http://mimi.case.edu/ontologies/2009/1/UnitsOntology#millimeter> ;
    dwc:measurementAccuracy "0.5"^^xsd:decimal ;
    dwc:measurementDeterminedDate "2009-08-22"^^xsd:date ;
    dwc:measurementDeterminedBy "Ryan B Stephens" ;
    dwciri:measurementDeterminedBy
         <http://scholar.google.com/citations?user=RAsUdjoAAAAJ> ;
    dwc:measurementMethod "unspecified" ;
    dwciri:measurementMethod <http://purl.obolibrary.org/obo/MMO_0000160> ;
    dwc:measurementRemarks "Accuracy from significant digits."@en].
```

Darwin Core
Resource
Relationship

DwC Resource Relationship

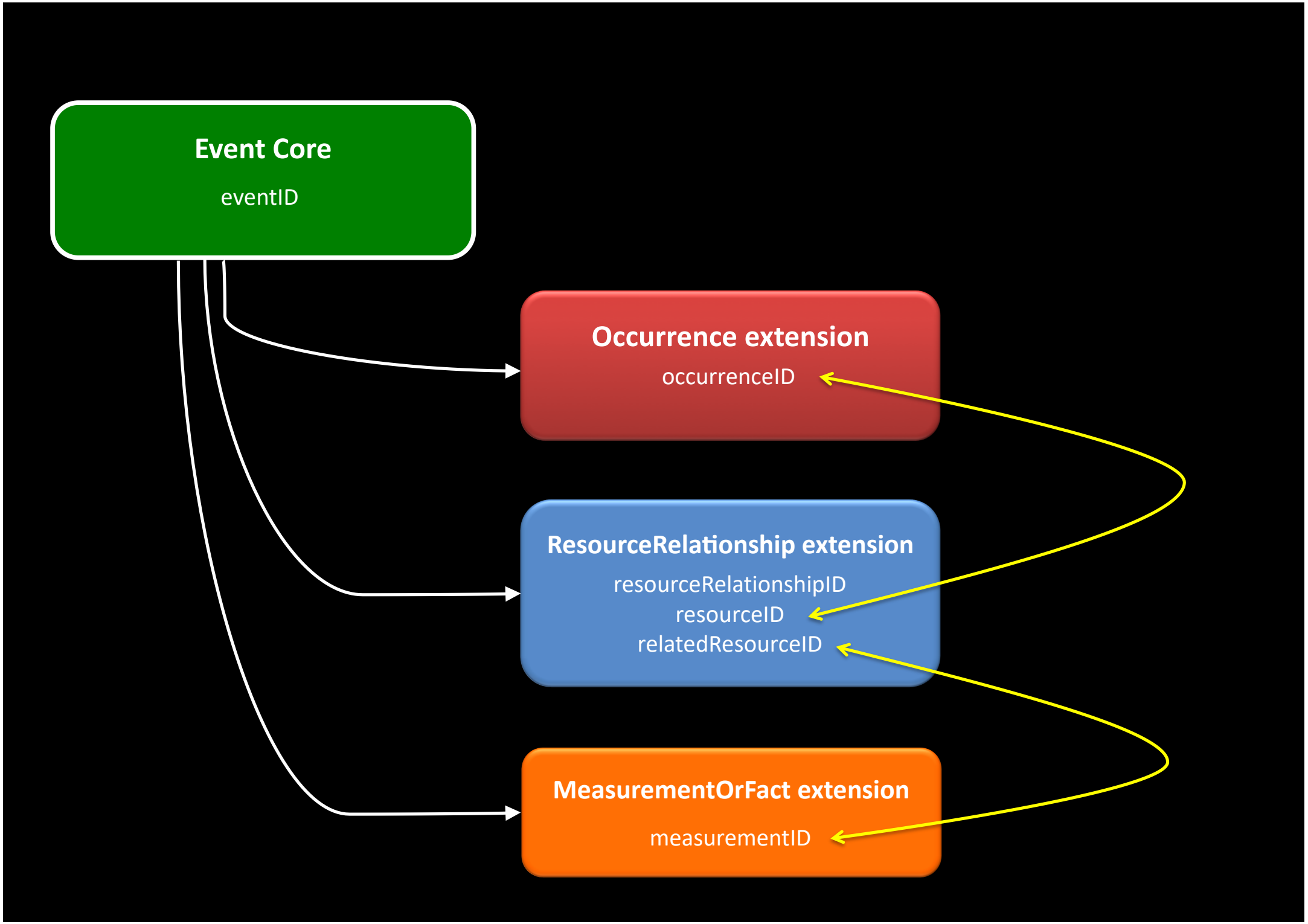
Term Name: ResourceRelationship

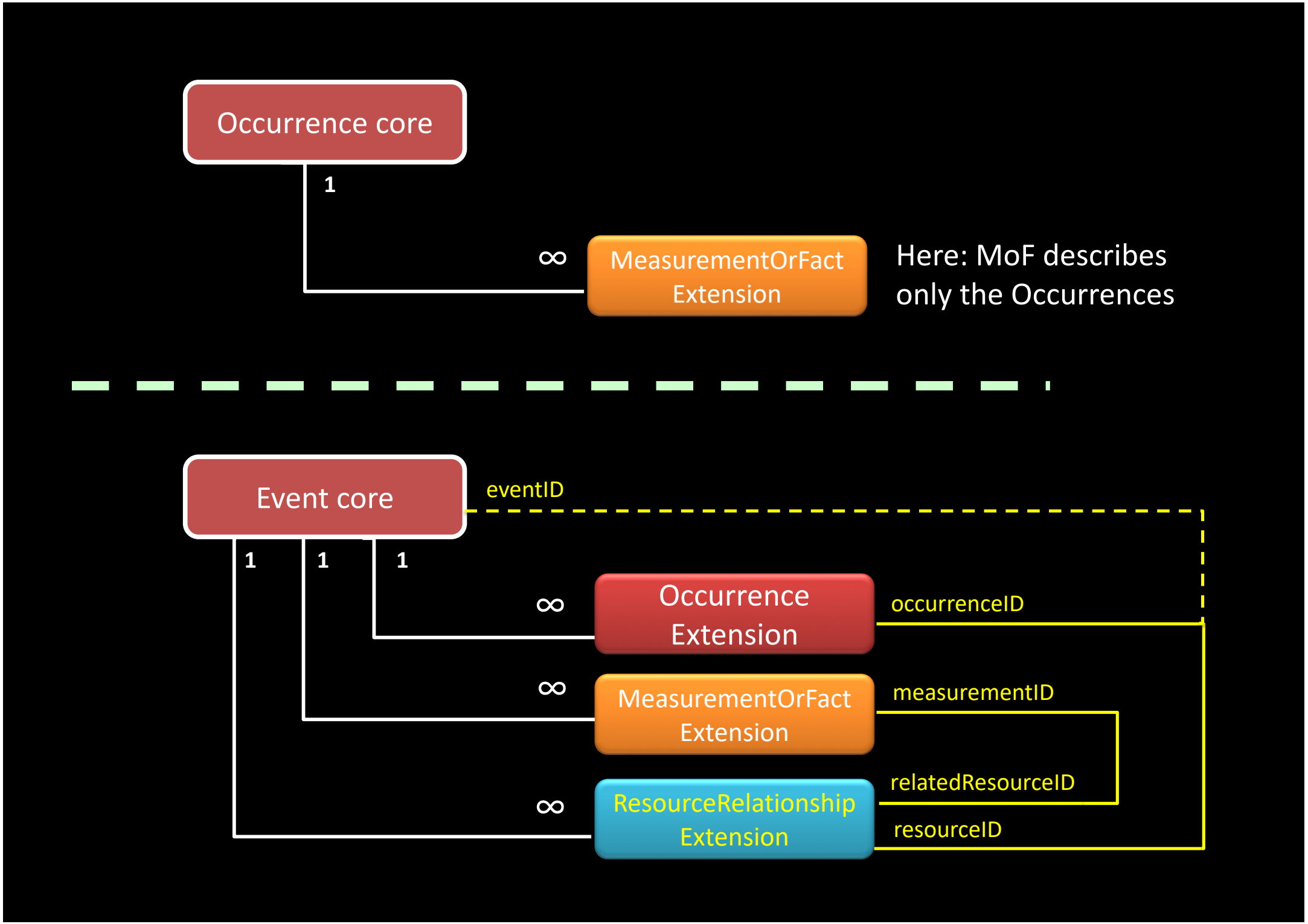
Identifier: http://rs.tdwg.org/dwc/terms/ResourceRelationship

Definition: A relationship of one rdfs:Resource (http://www.w3.org/2000/01/rdf-schema#Resource) to another.

Comment: Resources can be thought of as identifiable records or instances of classes and may include, but need not be limited to Occurrences, Organisms, MaterialSamples, Events, Locations, GeologicalContexts, Identifications, or Taxa. Example: An instance of an Organism is the mother of another instance of an Organism. For discussion see http://terms.tdwg.org/wiki/dwc:ResourceRelationship

ResourceRelationship | resourceRelationshipID | resourceID | relatedResourceID | relationshipOfResource | relationshipAccordingTo | relationshipEstablishedDate | relationshipRemarks


Event Core
eventID
Occurrence extension
occurrenceID
ResourceRelationship extension
resourceRelationshipID
resourceID
relatedResourceID
MeasurementOrFact extension
measurementID


Occurrence core
1
∞
MeasurementOrFact Extension
Here: MoF describes only the Occurrences
Event core
eventID
1
1
1
∞
Occurrence Extension
occurrenceID
∞
MeasurementOrFact Extension
measurementID
relatedResourceID
∞
ResourceRelationship Extension
resourceID

Use case for Resource Relationship

CORE entity is dwc:Event, measurement is on a dwc:Occurrence organized in a DwCA extension

- resourceRelationshipID = "relationship1"
- resourceID = "organism1"^^dwc:Occurrence
- relatedResourceID = "measurement1"
- relationshipOfResource = "measurement_of_occurrence"

Use case for Resource Relationship

CORE entity is dwc:Occurrence, but the measurement is on one of the entities flattened into the simple Darwin Core data record (Organism, Event, Location, Multimedia, MaterialSample, ...)

- resourceRelationshipID = "relationship1"
- resourceID = "organism1"^^dwc:Organism | "event1"^^dwc:Event | …
- relatedResourceID = "measurement1"
- relationshipOfResource = "measurement_of_organism/event ..."

Use case for Resource Relationship

CORE entity is dwc:Event (**measurement event**, not the collecting event), measurement is on a dwc:Occurrence published in another dataset

- resourceRelationshipID = "relationship1" | http:urn:uuid:UUID^^resourceRelationship
- resourceID = "http:urn:uuid:UUID"^^dwc:Occurrence
- relatedResourceID = http:urn:uuid:UUID^^dwc:MeasurementOrFact
- relationshipOfResource = "measurement_of_occurrence_in_another_dataset"

Proposing a new use case:

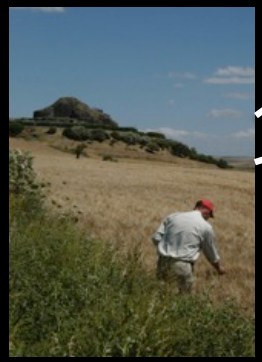
Measurement events as dwc:Event

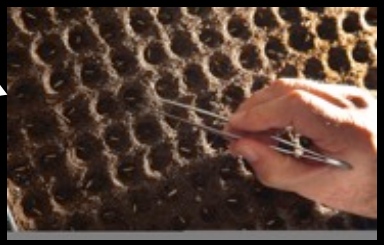
Experiments with living specimens at a location and date that is different from the **collecting event**.

Collecting expeditions

Genebank, collections

Field/lab experiments





Seeds from collecting trip expeditions are deposited in genebank collections.

Living seeds from genebank collections are used for field or laboratory experiments.

Germplasm material flow

- Material samples from many different collecting missions are deposited in different genebanks.
- Material from public genebanks are included in other genebank collections (and breeders collections).
- Material from different genebanks (and thus many different collecting events) are included in each field trial experiment.
- Collecting Event ≠ Measurement Event (MoF)

Use case
GBIF.no resolver

gbif.no resolver response

Event

urn:uuid:00b798d1-09cf-4571-b56a-06455b5de271

https://purl.org/gbifnorway/id/58c84deb-4d82-46e3-a109-527b2bac3d69.json

Occurrence

urn:uuid:58c84deb-4d82-46e3-a109-527b2bac3d69

https://purl.org/gbifnorway/id/58c84deb-4d82-46e3-a109-527b2bac3d69.n3

MeasurementOrFact

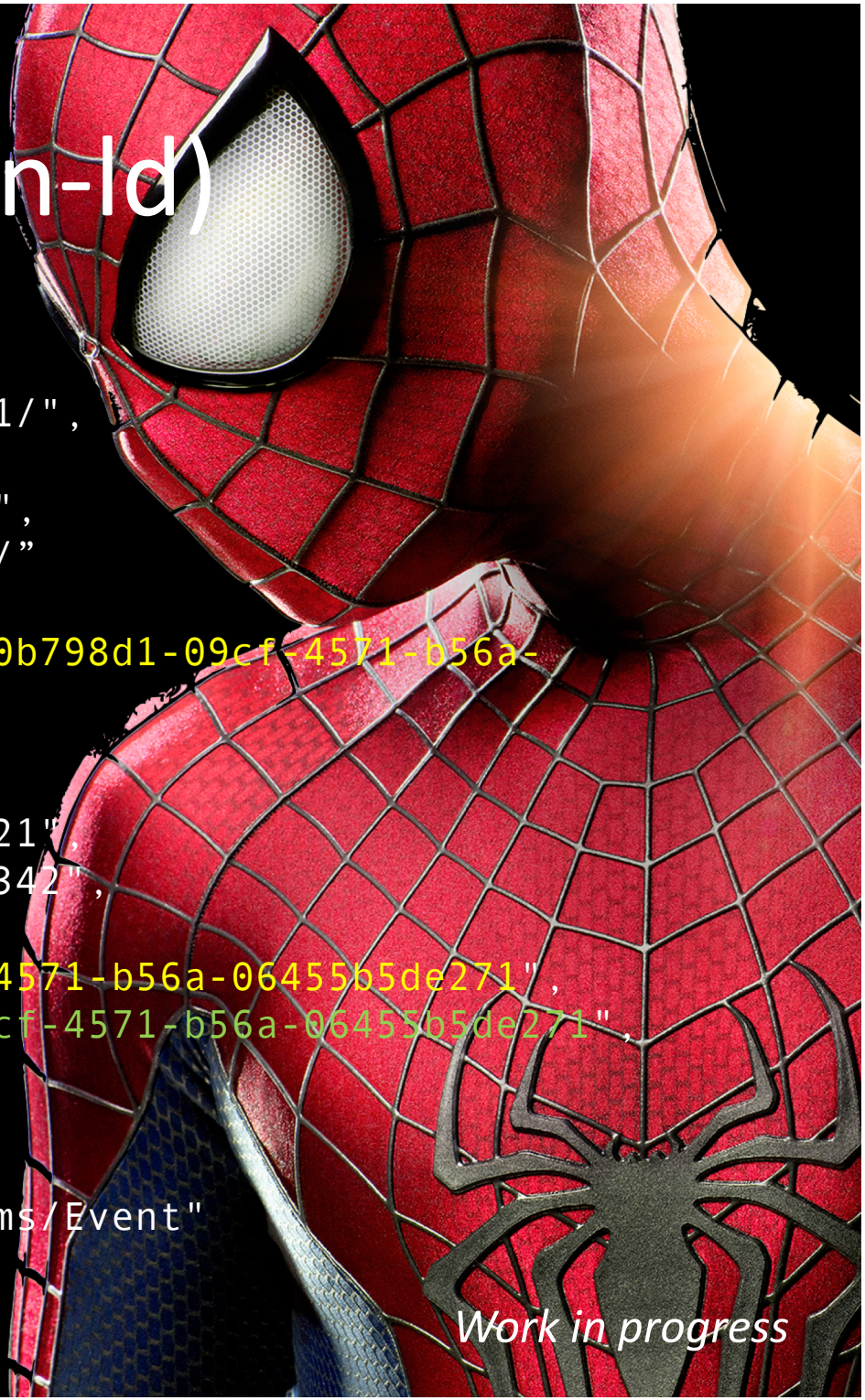
urn:uuid:5b0cd1c4-cf9e-4fbb-a995-f2b959d69f6c

https://purl.org/gbifnorway/id/5b0cd1c4-cf9e-4fbb-a995-f2b959d69f6c.rdf

https://data.gbif.no/resolver/

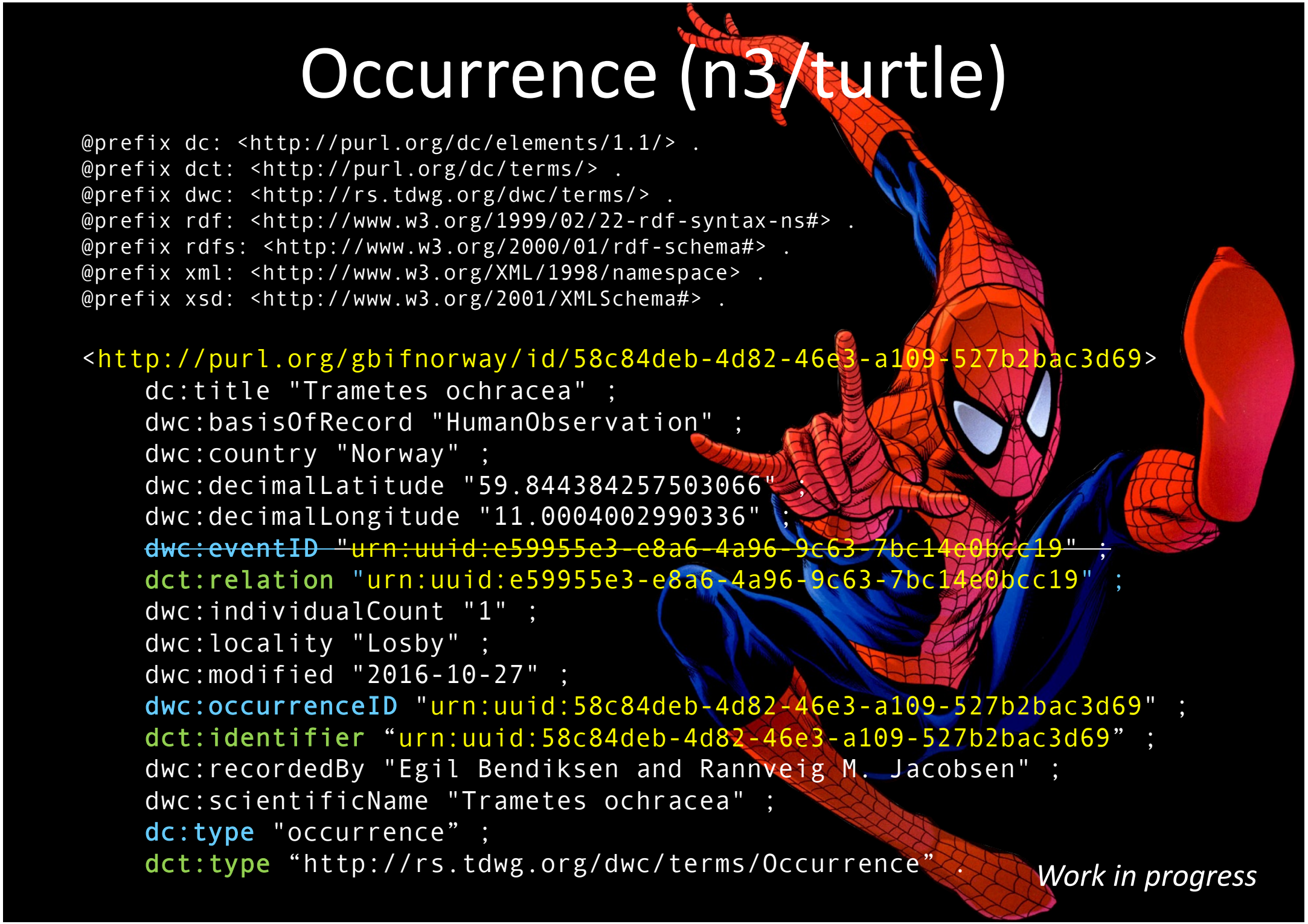
Event (json-ld)


```
{
  "@context": {
    "dc": "http://purl.org/dc/elements/1.1/",
    "dct": "http://purl.org/dc/terms/",
    "dwc": "http://rs.tdwg.org/dwc/terms/",
    ”dwciri”: “http://rs.tdwg.org/dwc/iri/”
  },
  "@id": "http://purl.org/gbifnorway/id/00b798d1-09cf-4571-b56a-
06455b5de271",
  "dc:title": null,
  "dwc:country": "Norway",
  "dwc:decimalLatitude": "60.10436984451321",
  "dwc:decimalLongitude": "10.58985049035342",
  "dwc:eventDate": "2013",
  "dwc:eventID": "urn:uuid:00b798d1-09cf-4571-b56a-06455b5de271",
  "dct:identifier": "urn:uuid:00b798d1-09cf-4571-b56a-06455b5de271",
  "dwc:locality": "L\u00f8venskiold",
  "dwc:modified": "2016-10-27",
  "dc:type": "event",
  "dct:type": "http://rs.tdwg.org/dwc/terms/Event"
}
```

Work in progress

Occurrence (n3/turtle)



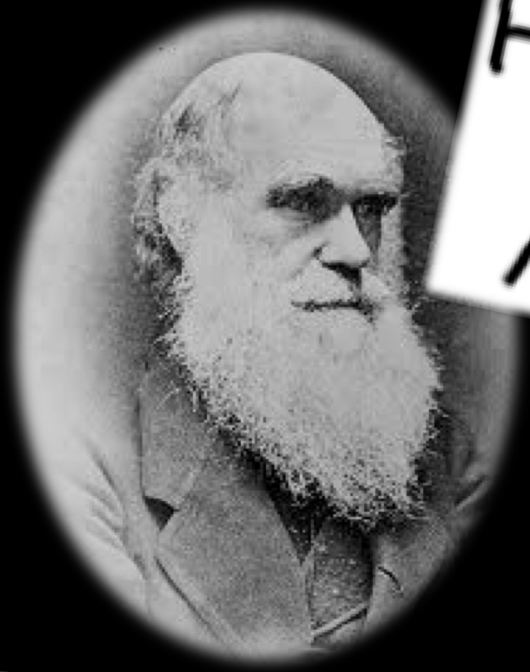
```
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dwc: <http://rs.tdwg.org/dwc/terms/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://purl.org/gbifnorway/id/58c84deb-4d82-46e3-a109-527b2bac3d69>
    dc:title "Trametes ochracea" ;
    dwc:basisOfRecord "HumanObservation" ;
    dwc:country "Norway" ;
    dwc:decimalLatitude "59.8443842575030 66" ;
    dwc:decimalLongitude "11.0004002990336" ;
    dwc:eventID "urn:uuid:e59955e3-e8a6-4a96-9c63-7bc14e0bcc19" ;
    dct:relation "urn:uuid:e59955e3-e8a6-4a96-9c63-7bc14e0bcc19" ;
    dwc:individualCount "1" ;
    dwc:locality "Losby" ;
    dwc:modified "2016-10-27" ;
    dwc:occurrenceID "urn:uuid:58c84deb-4d82-46e3-a109-527b2bac3d69" ;
    dct:identifier “urn:uuid:58c84deb-4d82-46e3-a109-527b2bac3d69” ;
    dwc:recordedBy "Egil Bendiksen and Rannveig M. Jacobsen" ;
    dwc:scientificName "Trametes ochracea" ;
    dc:type "occurrence” ;
    dct:type “http://rs.tdwg.org/dwc/terms/Occurrence” .
```

Work in progress

MeasurementOrFact (rdf)

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
   xmlns:dwc="http://rs.tdwg.org/dwc/terms/"
   xmlns:dc="http://purl.org/dc/elements/1.1/
   xmlns:dct="http://purl.org/dc/terms/"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
>
  <rdf:Description rdf:about="http://purl.org/gbifnorway/id/5b0cd1c4-
    cf9e-4fbb-a995-f2b959d69f6c">
    <dc:type>measurement</dc:type>
    <dct:type>http://rs.tdwg.org/dwc/terms/MeasurementOrFact</dct:type>
    <dwc:measurementType>Weight</dwc:measurementType>
    <dwc:measurementValue>25.0</dwc:measurementValue>
    <dwc:measurementUnit>gram</dwc:measurementUnit>
    <dwc:measurementID>urn:uuid:5b0cd1c4-cf9e-4fbb-a995-
f2b959d69f6c</dwc:measurementID>
    <dct:identifier>urn:uuid:5b0cd1c4-cf9e-4fbb-a995-
f2b959d69f6c</dct:identifier>
    <dwc:occurrenceID>urn:uuid:3686b360-62a3-41e9-85c3-
5b7bfbc57bed</dwc:occurrenceID>
 <dct:relation>https:purl.org/gbifnorway/id73686b360-62a3-41e9-85c3-
5b7bfbc57bed</dct:relation>
  </rdf:Description>
</rdf:RDF>
```

Work in progress



01001000 01101001 00100001

Slides available at https://bit.ly/gbifEu2018_new_datatypes