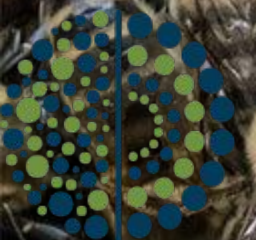


TDWG & DiSSCo

Mapping data standards for MIDS



Biodiversity
Information
Standards

TDWG

DiSSCo

Distributed System of Scientific Collections

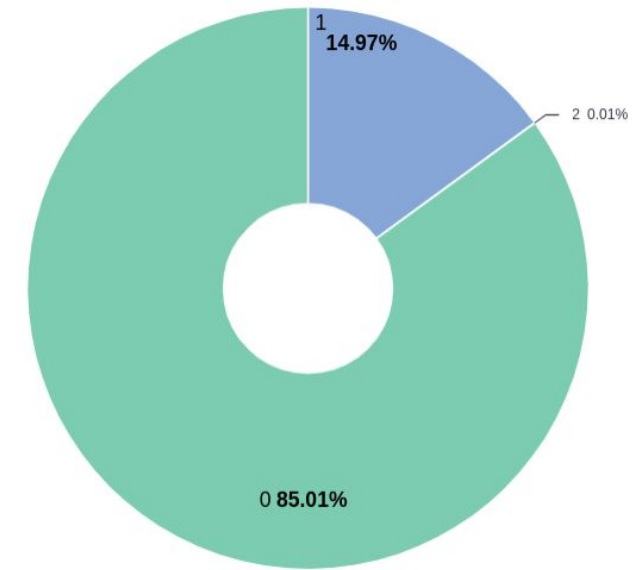
Sam Leeflang

Lead Programmer, Naturalis Biodiversity Center

FAIR-IMPACT, Nov 24, 2023

MIDS overview

- Minimal Information about a Digital Specimen
- A [Task Group](#) within the [Biodiversity Information Standards](#) (TDWG organization)
 - Under the [Collection Description Interest Group](#)
 - Led by Elspeth Haston (RBGE) and Cat Chapman (NAU & iDigBio)
- Aims are:
 - to provide a standard for measuring and monitoring digitisation
 - to provide a guide for prioritisation for digitisation
 - to align with existing mass digitisation programmes
 - to provide the mapping to help implementation
 - to provide recommendations to support data quality



The Information Element

- MIDS has Information Elements
 - Aimed at data completeness, not data quality
 - A specimen needs to have a list of attributes to comply
- MIDS is not a data standard is itself
 - It is not an ontology
 - It needs a mapping against multiple data standards
- Essential that every MIDS calculation uses the same mapping
 - MIDS will be calculated at multiple infrastructure levels

Name

SpecimenType

ObjectType

Modified

License



Bombus terrestris (Linnaeus, 1758), by Naturalis Biodiversity Center, licensed under CC0 1.0

Proposed solution

- Create a Simple Standard for Sharing Ontological Mappings (SSSOM)
 - Initially using the Simple Knowledge Organization System (SKOS)
- Create a mapping for each MIDS information element to other standards
 - Focus on the main data standards in the Biodiversity realm
 - Over 200 individual mappings against 9 different standard
- Multiple agents have started separately, and we later combined the result
 - Mathias Dillen (BR), Elspeth Haston (RBGE) and Sam Leeflang (Naturalis & DiSSCo)



| | | | | | | | | |
|-------------------------|-------|--------------------|--------------|---|-------------|---|--|------------------------|
| mids:PhysicalSpecimenID | mids0 | PhysicalSpecimenID | rdf property | http://www.w3.org/2004/02/skos/core#narrowMatch | narrowMatch | http://rs.tdwg.org/dwc/terms/materialSampleID | | material Sample ID |
| mids:PhysicalSpecimenID | mids0 | PhysicalSpecimenID | rdf property | http://www.w3.org/2004/02/skos/core#narrowMatch | narrowMatch | http://rs.tdwg.org/dwc/terms/catalogNumber | | catalog Number |
| mids:PhysicalSpecimenID | mids0 | PhysicalSpecimenID | rdf property | http://www.w3.org/2004/02/skos/core#narrowMatch | narrowMatch | http://rs.tdwg.org/dwc/terms/otherCatalogNumbers | | other Catalog Numbers |
| mids:PhysicalSpecimenID | mids0 | PhysicalSpecimenID | rdf property | http://www.w3.org/2004/02/skos/core#narrowMatch | narrowMatch | http://rs.tdwg.org/dwc/terms/occurrenceID | | occurrence ID |
| mids:PhysicalSpecimenID | mids0 | PhysicalSpecimenID | rdf property | http://www.w3.org/2004/02/skos/core#narrowMatch | narrowMatch | http://rs.tdwg.org/dwc/terms/materialEntityID | | material Entity ID |
| mids:Organization | mids0 | Organization | rdf property | http://www.w3.org/2004/02/skos/core#narrowMatch | narrowMatch | http://rs.tdwg.org/dwc/terms/ownerInstitutionCode | | owner Institution Code |
| mids:Organization | mids0 | Organization | rdf property | http://www.w3.org/2004/02/skos/core#narrowMatch | narrowMatch | http://rs.tdwg.org/dwc/terms/institutionID | | institution ID |
| mids:Organization | mids0 | Organization | rdf property | http://www.w3.org/2004/02/skos/core#narrowMatch | narrowMatch | http://rs.tdwg.org/dwc/terms/institutionCode | | institution Code |

Mapping Challenges

- The SSSOM is not a mapping between two data standards
 - MIDS is not an ontology but a list of elements
- Things get tricky when one MIDS element requires multiple data elements
 - For example: Quantitive Location needs the presence of three different data standard elements
 - They have an AND relationship
 - Most other elements have an OR relationship in that if any one of the data standard element is present, it fulfils the MIDS Information Element
- Looking at different ontologies for indication relationships
 - The W3C Web Ontology Language (OWL) -> intersectionOf to indicate AND relationships

| | | | | | | | | |
|---------------------------|-------|----------------------|--------------|---|----------------|---|--|-------------------|
| mids:QuantitativeLocation | mids2 | QuantitativeLocation | rdf property | https://www.w3.org/TR/2004/REC-owl-semantic-20040210/#owl_intersectionOf | intersectionOf | http://rs.tdwg.org/dwc/terms/decimalLatitude | | decimal Latitude |
| mids:QuantitativeLocation | mids2 | QuantitativeLocation | rdf property | https://www.w3.org/TR/2004/REC-owl-semantic-20040210/#owl_intersectionOf | intersectionOf | http://rs.tdwg.org/dwc/terms/decimalLongitude | | decimal Longitude |
| mids:QuantitativeLocation | mids2 | QuantitativeLocation | rdf property | https://www.w3.org/TR/2004/REC-owl-semantic-20040210/#owl_intersectionOf | intersectionOf | http://rs.tdwg.org/dwc/terms/geodeticDatum | | geodetic Datum |
| | | | | https://www.w3.org/TR/2004/REC-owl-semantic-20040210/#owl_intersectionOf | intersectionOf | http://rs.tdwg.org/dwc/terms/decimalLatitude | | decimal Latitude |
| | | | | | | http://rs.tdwg.org/dwc/terms/decimalLatitude http://rs.tdwg.org/dwc/terms/decimalLongitude http://rs.tdwg.org/dwc/terms/coordinateUncertaintyInMeters http://rs.tdwg.org/dwc/terms/coordinatePrecision http://rs.tdwg.org/dwc/terms/geodeticDatum | | |

Questions

- Are we using SSSOM correctly by mapping information elements against data standard elements?
- What would be the best way to indicate OR and AND relationships over multiple data elements?
- How would the mapping perform against standards with different serialisation formats?
- What are the best practices for both machine use and human use of the mapping?





Huge thanks to the whole TDWG MIDS Task Group
Elspeth Haston and Mathias Dillen in particular

THANK YOU FOR YOUR ATTENTION !



Biodiversity
Information
Standards
TDWG

