An Ontology-based Reference Process to Provide Interoperability and Foster Database Integration for Biodiversity Data

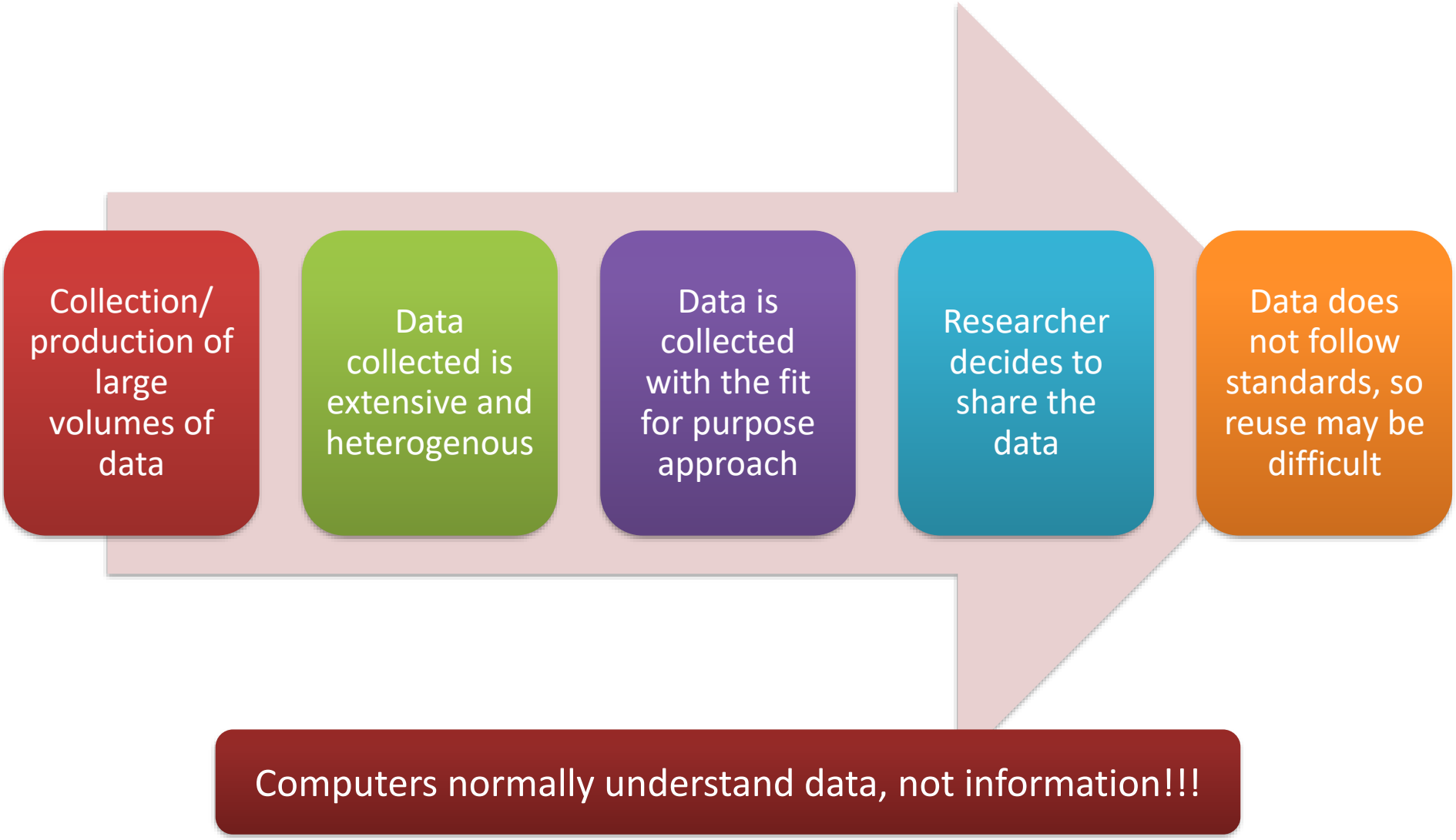
Fabiana Santana, Silvia Scheunemann,
Anarosa Brandão & Antonio Mauro Saraiva
Fabiana.Santana@canberra.edu.au

Financial support: FAPESP Fundação de Amparo à Pesquisa do Estado de São Paulo

Project: openModeller

Collaboration: University of Canberra, POLI USP, bccvl

The Problem


Collection/ production of large volumes of data
Data collected is extensive and heterogenous
Data is collected with the fit for purpose approach
Researcher decides to share the data
Data does not follow standards, so reuse may be difficult
Computers normally understand data, not information!!!

Why Sharing?

Reproducibility

Provenance

Continuity

Data Reuse

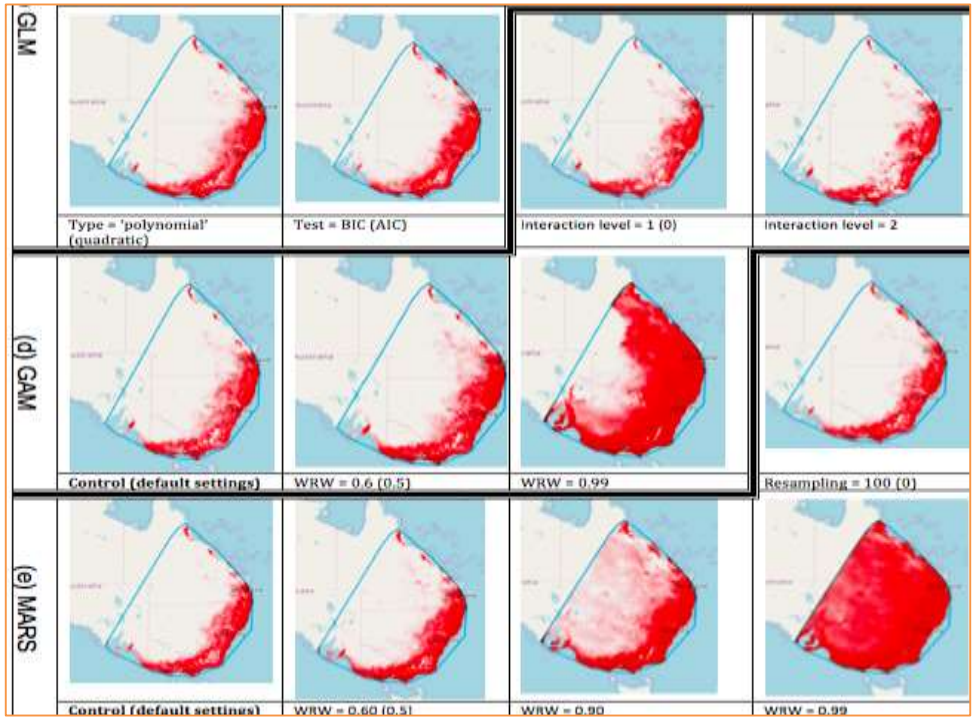
Assessment of Contributions

Ultimately: Science Evolution!

I’m normally suspicious of research that cannot be reproduced or does not have the basic provenance information…

Why?

- Example – Koala's Species Distribution Modelling:



- Sharing the data should be a requirement for publication!

How?

- Better to start using the standards even if you do not plan on sharing the data!
  - Easy: just name the columns properly;
  - Guide on what kind of information may be associated with your domain of study;
  - May help other researchers in the future.

How?

- And what about the data that has already been collected?
  - CSV file;
  - Microsoft (R) Office Excel (R) spreadsheet or similar;
  - Paper.

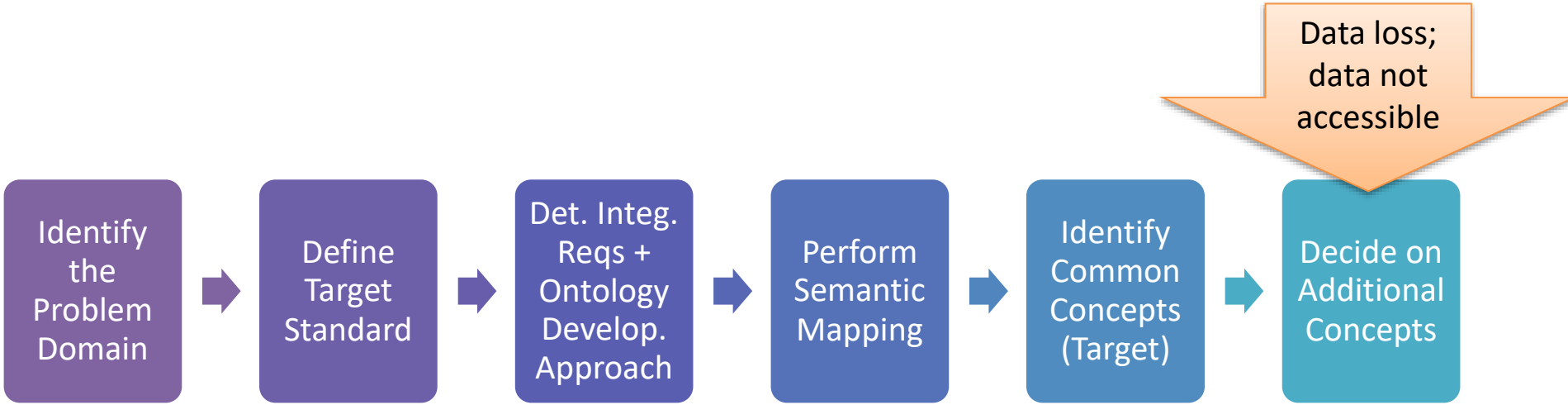
- Can this data be shared?

How?

- The proposed solution:
  - It's a reference process:
    - Can be adapted to potentially any domain;
  - Purposes:
    - Interoperability;
    - Data sharing; and
    - Database integration of biodiversity data.

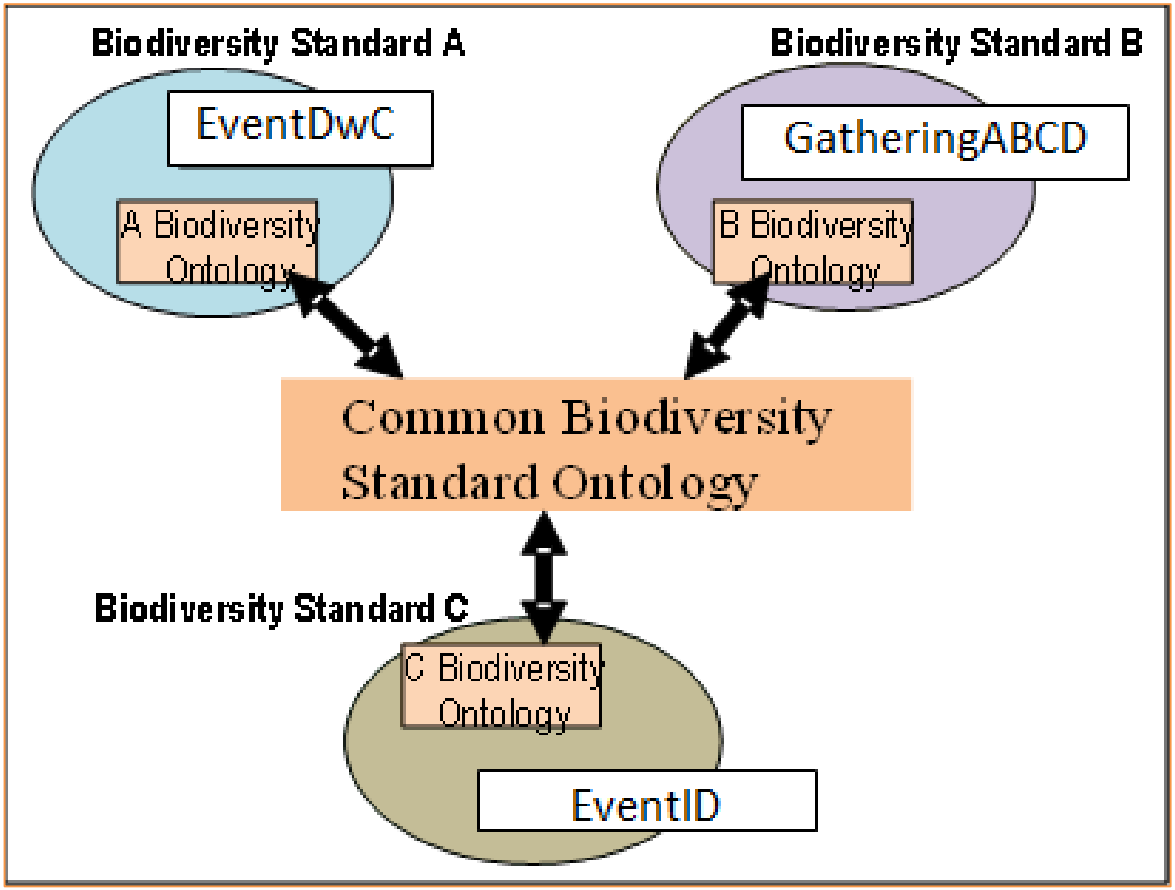
Reference Process

- Based on Ontologies:
  - Powerful technique to provide interoperability among datasets and information systems;
  - Mainly focused on the semantic analysis of existing data.



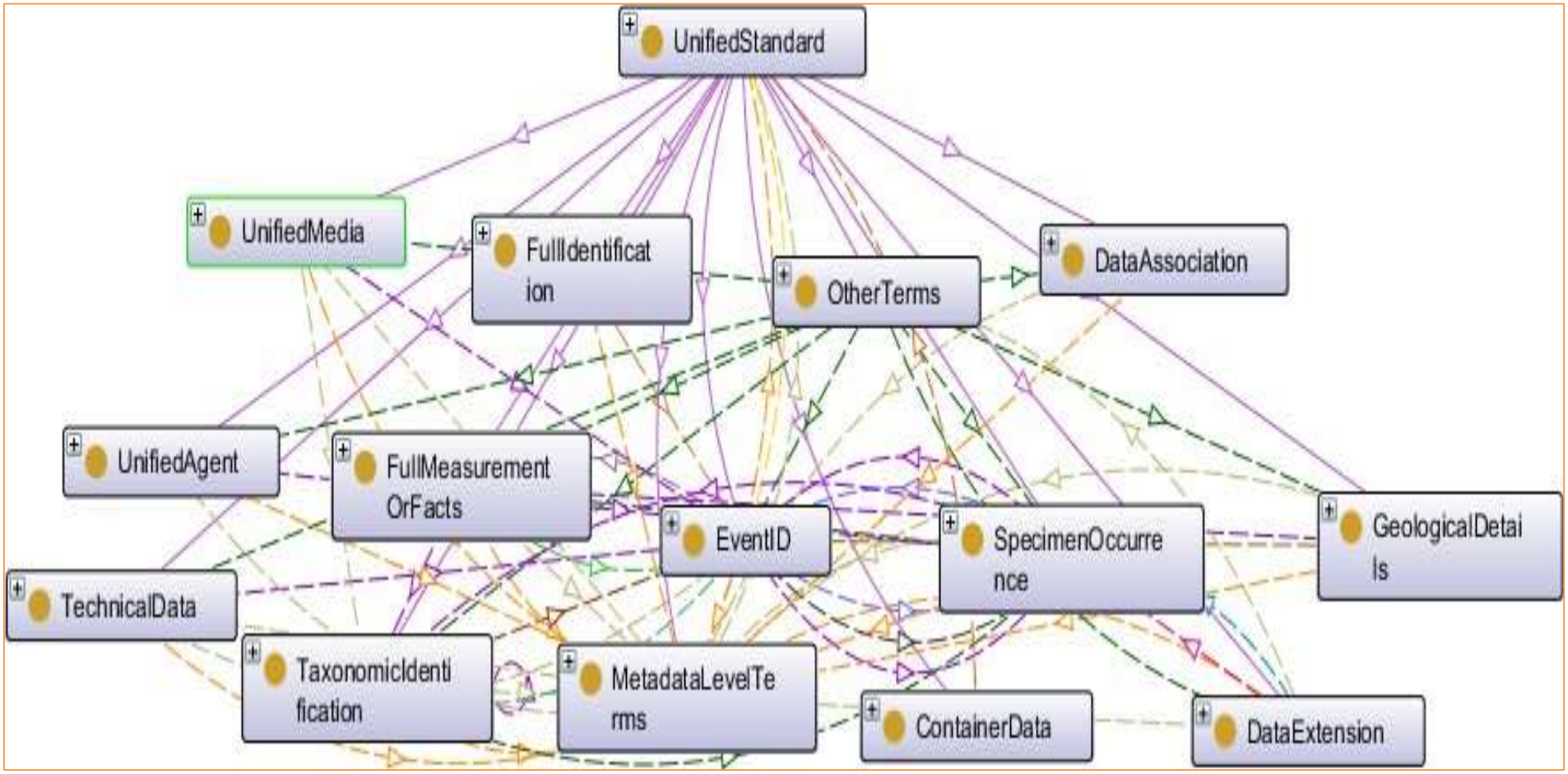
Case Study

- *FOBioS: Functional Ontology for Biodiversity Standards*
  - Proposed for the integration of datasets based on ABCD and Darwin Core.



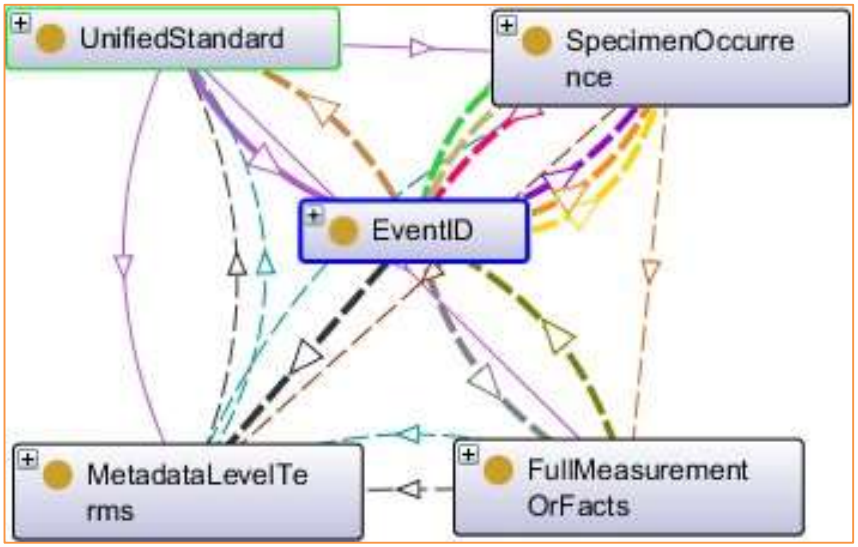
Ontology

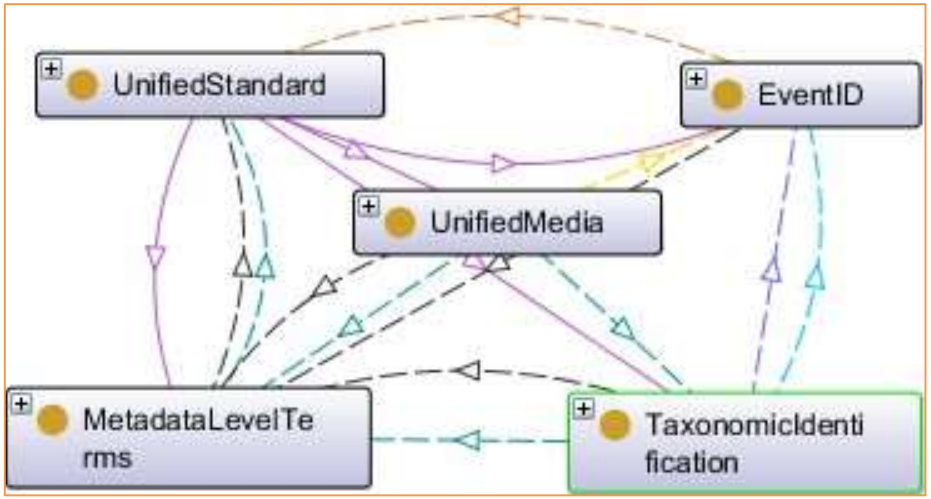
- *FOBioS* Overview



Ontology

- *FOBioS* fully define each of the relevant concepts in the Ontology:
  - *EventID, TaxonomicIdentification, GeologicalDetails, UnifiedMedia, UnifiedAgent, etc.*
  - Examples – *EventID* & *UnifiedMedia*

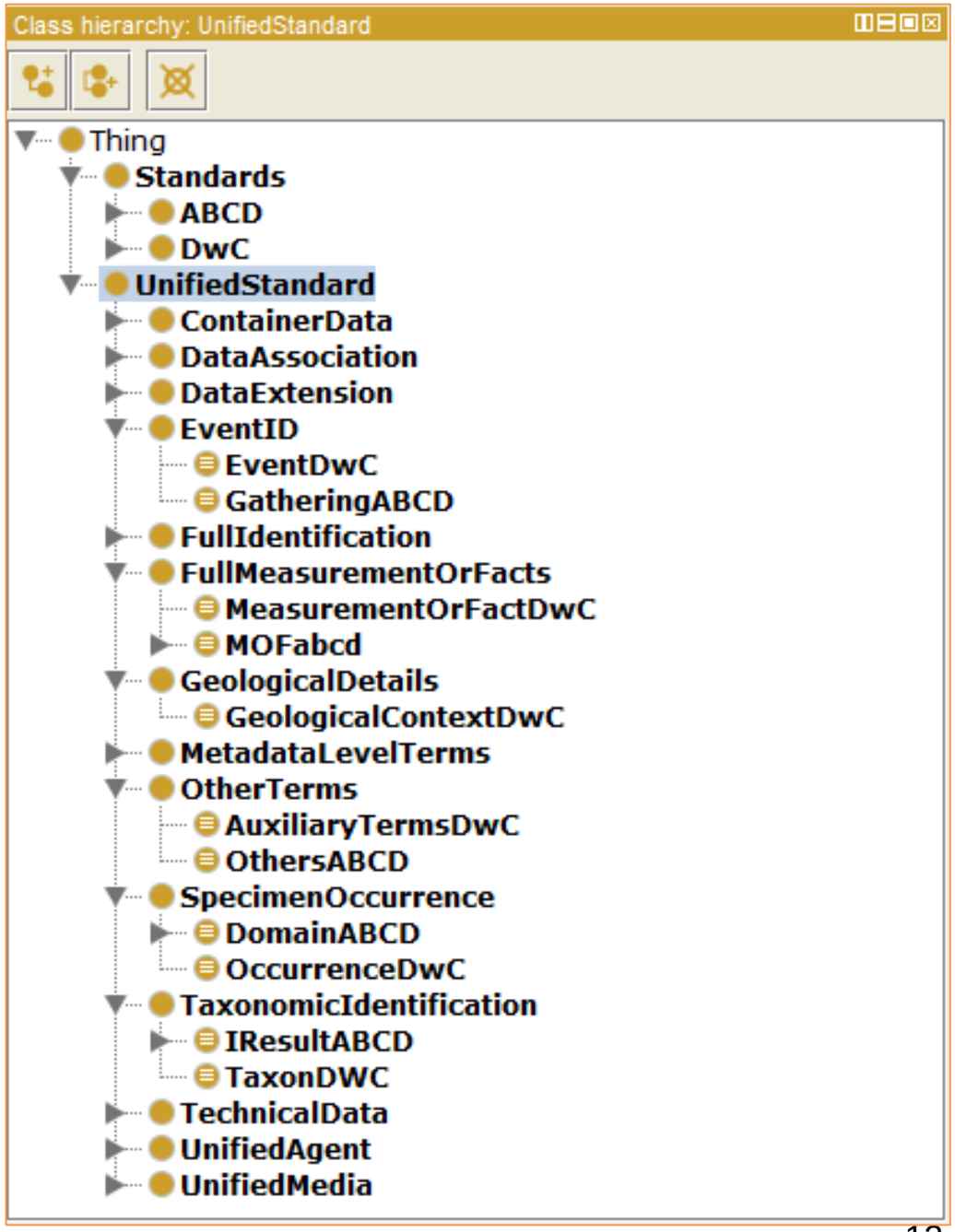




protégé

- *protégé* is an “open-source ontology editor and framework for building intelligent systems” [protege.stanford.edu];

- Used in this work to create and test the *FOBioS* ontology.



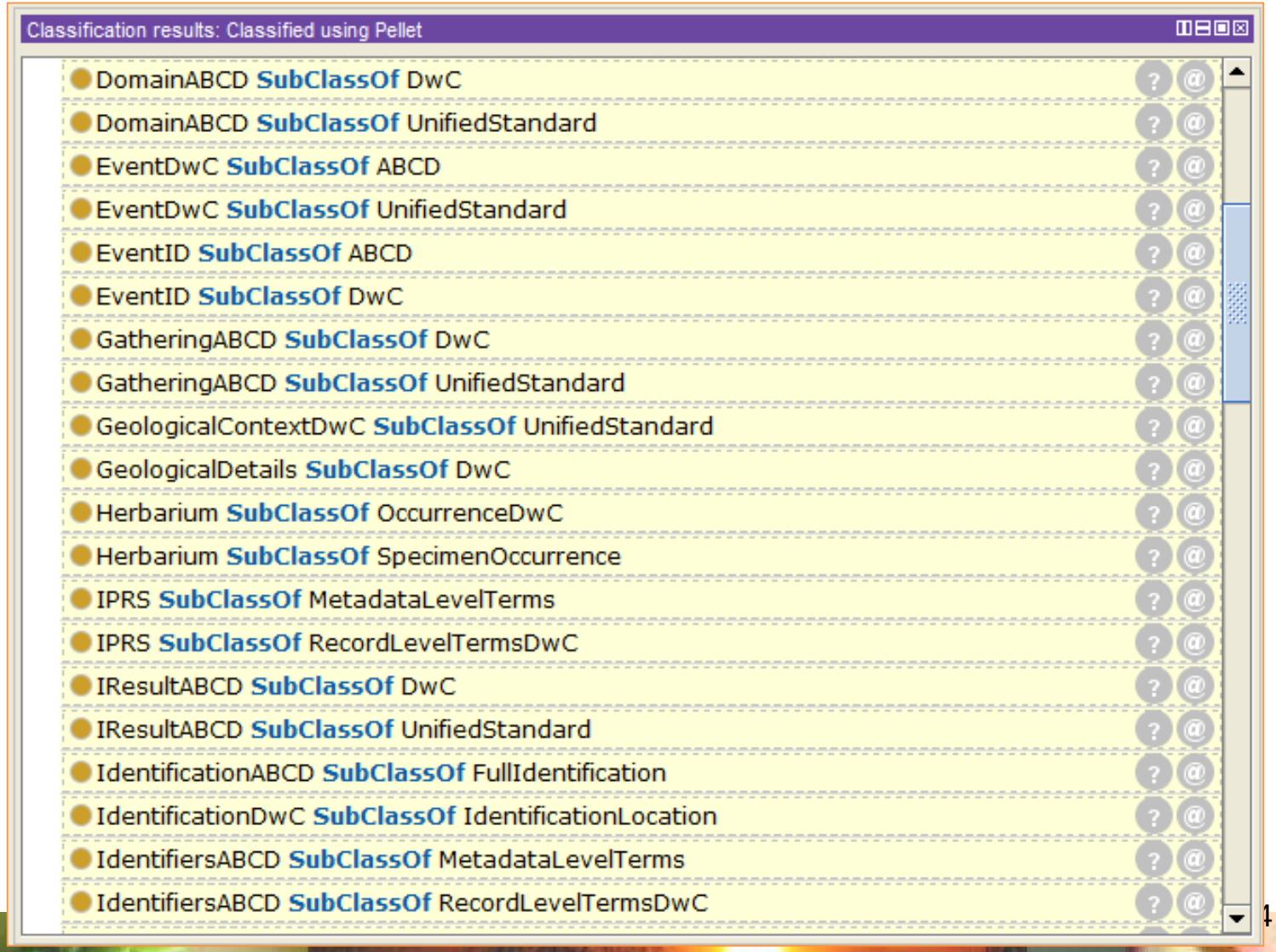
Mapping - Ontology Concepts

FOBiOS	ABCD
EventID	GatheringABCD
SpecimenOccurrence	DomainABCD
TaxonomicIdentification	IresultABCD
GeologicalDetails	
FullMeasurementOrFacts	MOFabcd
UnifiedMedia	MediaABCD
UnifiedAgent	AgentABCD
TechnicalData	TechnicalABCD
OtherTerms	OtherABCD
MetadataLevelTerms	MetaABCD
DataExtension	UnitExtensionABCD
DataAssociation	AssociationAssemblageABCD
ContainerData	ContainerABCD
FullIdentification	IdentificationABCD

FOBiOS	Darwin Core
EventID	EventDwC
SpecimenOccurrence	OccurrenceDwC
TaxonomicIdentification	TaxonDwC
GeologicalDetails	GeologicalContextDwC
FullMeasurementOrFacts	MeasurementOrFactsDwC
UnifiedMedia	
UnifiedAgent	
TechnicalData	
OtherTerms	AuxiliaryTermsDwC
MetadataLevelTerms	RecordLevelTerms
DataExtension	
DataAssociation	
ContainerData	
FullIdentification	LocationDwC
	IdentificationDwC

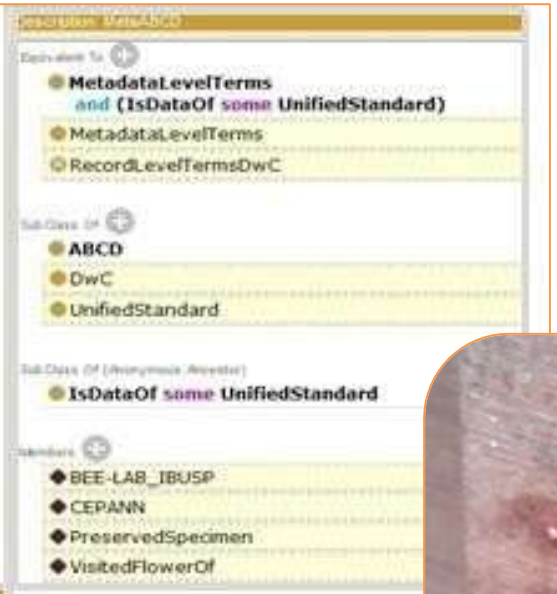
Inferences on *protégé*

- Compatible with mapping:



FOBioS: species Lestrimelitta limao

Padrão Darwin Core



Padrão ABCD

FOBiOS

Source: Jonasfleith - https://commons.wikimedia.org/w/index.php?curid=31487194

Conclusion

- The reference process can help to automate data sharing/database integration by providing interoperability between standards/different semantic approaches to data storing;

- Foster research improvement by supporting improvements in:
  - Reproducibility;
  - Provenance;
  - Continuity;
  - Data Reuse; and
  - Assessment of Contributions.

An Ontology-based Reference Process to Provide Interoperability and Foster Database Integration for Biodiversity Data

Fabiana Santana, Silvia Scheunemann, Anarosa Brandão & Antonio Mauro Saraiva

Fabiana.Santana@canberra.edu.au

ACKNOWLEDGMENT

The authors are grateful to Fapesp and to the Faculty of Science and Technology, University of Canberra, for the financial support.

FAPESP
Fundação de Amparo à Pesquisa do Estado de São Paulo

UNIVERSITY OF CANBERRA