A knowledge-based system for discovering ecological interactions

in biodiversity data-stores of heterogeneous specimen-records:

A case-study of flower-visiting ecology

Willem Coetzera,b, Deshendran Moodleyb, Aurona Gerberc

Coetzer, W., Moodley, D., Gerber, A., A knowledge-based system for discovering ecological interactions in biodiversity data-stores of heterogeneous specimen-records: A case-study of flower-visiting ecology, *Ecological Informatics* (2014), DOI: 10.1016/j.ecoinf.2014.06.

aSouth African Institute for Aquatic Biodiversity, Private Bag 1015, Grahamstown 6140, South Africa

bCAIR: Centre for Artificial Intelligence Research at the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag X54001, Durban 4000,South Africa

cCAIR: Centre for Artificial Intelligence Research at the CSIR Meraka Institute, P.O. Box 395, Pretoria   
0001, South Africa

[w.coetzer@saiab.ac.za](mailto:w.coetzer@saiab.ac.za) (Willem Coetzer, corresponding author), Tel. +27 46 603 5841

[moodleyd37@ukzn.ac.za](mailto:moodleyd37@ukzn.ac.za) (Deshendran Moodley), Tel. +27 12 841 2422

[agerber@csir.co.za](mailto:agerber@csir.co.za) (Aurona Gerber), Tel. +27 31 260 1019

**Highlights**

* We model expert knowledge of flower-visiting behavioral ecology in an OWL ontology.
* The perspective of the ontology is event-centric.
* This perspective facilitates the extraction of ecological interactions.
* We describe a system which transforms specimen-data into ecological interactions.
* We evaluate the system’s scalability, extension and potential impact.

**Abstract**

We modeled expert knowledge of arthropod flower-visiting behavioral ecology and represented this in an event-centric domain ontology, which we describe along with the ontology construction process. Two smaller domain ontologies were created to represent expert knowledge of known flower-visiting insect groups and expert knowledge of the flower-visiting behavioral ecology of *Rediviva* bees. Two application ontologies were designed, which, together with the domain ontologies, constituted the ontology framework of a prototype semantic enrichment and mediation system that we designed and implemented to improve semantic interoperability between flower-visiting data-stores. We describe and evaluate the system implementation in a case-study of three flower-visiting data-stores, and we discuss the system’s scalability, extension and potential impact. We demonstrate how the system is able to dynamically extract complex ecological interactions from heterogeneous specimen data-stores. The conceptual stance and modeling approach are potentially of general use in representing knowledge of animal behavior and ecological interactions, and in engineering semantic interoperability between data-stores containing behavioral ecology data.

Keywords: Semantic interoperability; biodiversity and ecosystem informatics; flower-visiting behavioral ecology; data integration; ontology; conceptual modeling; event-centric perspective

1. **Introduction**

Behavior and ecological interactions, between individual organisms and between species, are hallmarks of biodiversity, distinguishing biodiversity from the subjects of other natural sciences such as geology or chemistry. It is often the complexity, variability, patterns, and importance of behavior and ecological interactions that motivate biodiversity scientists and ecologists to study biodiversity in the applied context of agriculture (e.g. pest control) or conservation (e.g. invasive species). The most important method of collecting data on behavior or ecological interactions is by directly observing animals (i.e. the field of ethology and the field of behavioral ecology: Krebs & Davies, 1996), though there is much interest in developing technologies to enable remote, automated biodiversity observation in the sense of Earth observation (Collins *et al*., 2006; Hart & Huang, 2012; Scholes *et al*., 2008).

In the field of Biodiversity and Ecosystem Informatics (BDEI) scientists typically analyze data which originate from specimen collections, usually held by natural history museums. These descriptive ‘specimen data’ have received much attention in the sense of fitness-for-use (e.g. completeness, accuracy and precision) (Bisby, 2000), and the focus is now turning to the meaning of biodiversity information. Analyzing ecological interactions is considered a priority in BDEI (Peterson *et al*., 2010) because biodiversity scientists need not only descriptive knowledge but explanatory knowledge of biodiversity and ecological processes that will be useful to Society. Generally in BDEI there is a need to improve semantic interoperability between biodiversity data-stores, to more easily and meaningfully aggregate data and mine the aggregations for useful biodiversity information.

In this paper we describe and evaluate a prototype semantic enrichment and mediation system for improved semantic interoperability between three museum data-stores containing specimen-records of flower-visiting insects, including annotations of insect flower-visiting behavior and behavioral ecology. We found that flower-visiting ecologists have expert or implicit knowledge of ecological interactions that is partially and differently expressed in, but may be missing from, the specimen data. This forces the experts themselves to manipulate the data using manual techniques that are neither consistent nor efficient and which do not transform the data into information, meaning that the output of the analysis is still specimen data. The mediation system we developed, however, was able to use knowledge of behavioral ecology, represented in ontologies, to consistently transform biodiversity specimen data into useful ecological information emphasizing ecological interactions.

Whereas ontology modeling has furthered the representation of general concepts used in specimen collections in natural history museums (Baskauf & Webb, 2014; Walls *et al*., 2014; Wieczorek *et al*., 2012), there is a need to model specific concepts that characterize the diversity and uniqueness of particular groups of phylogenetically or ecologically related species, specifically behaviors that represent ecological interactions between individuals and between species, such as pollination, parasitism and predation. Examples of such groups or behaviors include dolphins, wasps, insects that visit flowers, the pests of stored grain, or fish that swim past telemetry stations. In this paper we highlight a typical case of biodiversity uniqueness in the behavior and ecological interactions between plants and the specialized oil-collecting *Rediviva* bees of southern Africa. This is a special case of the general theme of flower-visiting by arthropods, and it exemplifies the local ‘variation on a theme’ that is typical of biodiversity and behavioral ecology.

We observed that the utilization of biodiversity data from specimen collections, including from natural history museums, can be overly data-centric because museum scientists (e.g. taxonomists and systematists) tend to focus on specimens and their attributes. This may also be true of the utilization of data collected during ecological surveys that do not yield physical specimens needing curation, but which nevertheless emphasize the importance of the occurrence records (also called ‘observations’) and their spatial and temporal attributes. This has been referred to as the ‘what-where-when approach’. Unless a sampling protocol is specifically designed for collecting behavioral data (‘what was it doing?’) or ecological interaction data (‘how or why; and what was it doing that with, or to, or on?’), it can be difficult to extract meaningful ecological information from biodiversity data, especially data associated with specimens in natural history collections.

The need therefore arises to bridge this gap between specimen data and ecological information. In our case-study one way to do this was to view visits to flowers by arthropods from an event-centric perspective (Worboys, 2005). This affords a view of both behavior and ecological interactions (i.e. occurrents) from a level higher than that of the observer who sees the data as attributes of plant organisms and insect organisms (or continuants) which become preserved as specimens in natural history collections. Encoded on the specimen labels and in the database records documenting those labels are pieces of a puzzle that do not form a picture of a museum drawer containing pinned bees. Rather, the elements of the picture are the interactions between bees and plants, and the picture communicates the composition of the interactions and their relationships between themselves and with other things (e.g. predators) and events (e.g. heat waves). This view is more commensurable with the intention of an ecologist to acquire knowledge of the ecological relationships between arthropods and plants on a more general level, while looking down to the origin of much biodiversity information in specimen collections in natural history museums. In this paper we therefore hope to offer a more knowledge-centric solution that will give expression to the implicit knowledge of specialist ecologists who would otherwise be forced to use tools that reinforce a data-centric perspective.

Establishing cause-and-effect in the study of behavior or biotic interactions, however, requires expert or implicit knowledge of behavioral ecology to be represented explicitly. Moreover, scientists’ observations of behavior cannot be complete, yet they need to extract as much information from their observations as is possible. Even a complete ontological knowledge model is discrete and offers no way to assign a probability to an event. Ecologists typically study the effects of global change on ecosystems using interaction networks, among which Bayesian models are important (Aderhold *et al*., 2012). We therefore see the ultimate challenge as one of combining the expressivity of a knowledge model with the predictivity of a Bayesian model. This hybrid knowledge representation modeling approach has been used in the Earth Observation domain to detect wildfires (Moodley *et al*., 2012).

Our primary objective is to show that through semantic enrichment and mediation an ontology that represents expert or implicit knowledge can be used to transform traditional, heterogeneous biodiversity specimen data, containing detailed flower-visiting behavioral ecology annotations, into useful biodiversity or ecological information. Moreover, this enrichment and mediation can be automated. We further suggest that the automation can be employed in the construction of standardized flower-visiting networks that can be used to facilitate studies of flower-visiting in different contexts. In order to do all of the above, an information system needs to be capable of distinguishing between common biodiversity data elements and specific knowledge of flower-visiting behavioral ecology.

Our secondary objective is to convey the design of a generalized system architecture for semantic enrichment and mediation in behavioral ecology. We have studied a particular theme (flower-visiting) and a particular group of species (flower-visiting by *Rediviva* bees) but we propose that this generalized system architecture for biodiversity and behavioral ecology may be extended to other themes, groups of species and contexts. The ontological perspective on events could be an important way to reduce the complexity of representing expert knowledge of animal behavior and ecological interactions. Importantly, our approach to developing a conceptual model of behavior and behavioral ecology keeps an eye on how ontology classes can be practically linked to specimen-records or occurrence-records in biodiversity data-stores. Instead of modeling behavior or behavioral ecology to develop an expressive or precise ontology, our ontological framework has a specific utilitarian place and purpose in the architecture of an information system.

In Section 2 we refer to literature on semantic interoperability in BDEI to sketch the background, and in Section 3 we introduce our case-study of semantic mediation and interoperability between specimen-records originating from three natural history museums. In Section 4 we explain the process of ontology construction and describe our core domain ontology of arthropod flower-visiting behavioral ecology as well as two smaller domain ontologies. In Section 5 we describe the architecture and implementation of a prototype semantic enrichment and mediation system, and in Section 6 we evaluate the system implementation by considering how well the system automates the transformation of flower-visiting data into useful ecological information. In the system evaluation particular attention is given to the three kinds of semantic enrichment performed by the system and the assumptions inherent in each, the resolution of missing data, and the extraction of new information from the data. We discuss the potential impact of the implemented enrichment and mediation system in studies of flower-visiting arthropod ecology. In Section 7 we conclude this work and outline future work.

1. **Literature review and background**
   1. Semantic interoperability in BDEI

One of the agreed fundamental objectives of BDEI is to improve semantic interoperability between distributed, heterogeneous biodiversity data-stores (Deans *et al*., 2011; Edwards *et al*., 2000; Jones *et al*., 2006; Michener & Jones, 2012) through the use of semantic web technologies (Antezana *et al*., 2009; Daltio & Medeiros, 2008). The need for data aggregation arises from the localized uniqueness and wide geographic distribution of biodiversity; understanding the general spatio-temporal patterns in biodiversity usually requires datasets to be aggregated. The range of concepts in BDEI is extremely wide and deep (Madin *et al*., 2007). Interoperability—especially of the semantic kind—is lacking and needed because biodiversity data originate from so many communities and sources, and datasets are more often than not heterogeneously structured and they encode the same concepts that are (slightly) differently defined (e.g. ‘pollinator’ can have a broad or very specific meaning).

The need for semantic interoperability among different user communities has been articulated in oceanography (Graybeal *et al*., 2012), including among users of marine biodiversity data, and is well established in ecology (Madin *et al*., 2008; Michener & Jones, 2012; Michener *et al*., 2007; Michener *et al*., 2011). In oceanography and ecology there is much to gain from solving the problems of integration and interoperability between biotic data or systems and those that have an abiotic focus e.g. environmental sensor networks (Collins *et al*., 2006). Semantic interoperability is no less important in biological taxonomy (Deans *et al*., 2011), which has a long history and many specialized communities of practise that focus on specific groups or taxa (e.g. botanists, entomologists, mycologists and many others). In BDEI datasets typically encompass elements of all of the above domains as well as others, such as the socio-economic domain. For example pollination is both an ecologically important and an economically valuable ecosystem service (Gallai *et al*., 2009).

Ontologies can be used to enable semantic interoperability. The challenge of engineering semantic interoperability in BDEI using ontologies has been addressed conceptually (e.g. Michener & Jones, 2012; Michener *et al*., 2007). Few practical solutions have been implemented although an early example appeared in 2008 (Daltio & Medeiros, 2008). Ontologies have also been used more specifically in BDEI to link genotype to phenotype (Peterson *et al*., 2010) to discover patterns of gene expression (e.g. Cooper *et al*., 2013; Sala & Bergamaschi, 2009).

Ontology engineering in BDEI is relatively young. The emerging Darwin Semantic Web (DSW) ontology (Baskauf & Webb, 2014) contains classes originating from the Darwin Core set of terms (Wieczorek *et al*., 2012), which was among the first data standards for publishing and integrating biodiversity data. The Biological Collections Ontology (BCO), which complies with the Basic Formal Ontology (BFO), serves a general purpose similar to that of DSW, but covers a much broader range of use-cases in biodiversity informatics (including e.g. sampling processes) (Walls *et al*., 2014). DSW articulates the specific classes needed for expressing the concepts traditionally used when analysing specimen data from natural history collections. Both DSW and BCO are occurrence-centric with respect to classes that contain the entities of biodiversity. The Population and Community Ontology (PCO), also BFO-compliant, contains classes for representing ‘material entities, qualities, and processes related to collections of interacting organisms such as populations and communities’ (Walls *et al*., 2014). PCO therefore introduces classes that can be used to relate interacting biodiversity entities to each other. Together with the Environment Ontology, BCO and PCO potentially cover (Walls *et al*., 2014) the broad and deep range of classes needed for reasoning over biodiversity concepts in all their dimensions, at different levels of organization (e.g. genetic or ecological) and in the different contexts commonly encountered, including ecological surveys and natural history collections of physical specimens.

* 1. Ontology modeling in behavioral ecology

The use of ontology modeling in the study of behavior has been addressed in neurobiology (Gkoutos *et al*., 2012). The Neuro Behavior Ontology contains classes of two fundamental types, namely *BehavioralProcess* and *BehavioralPhenotype*, the sub-classes of which constitute a species-independent behavior vocabulary that is interoperable with the Gene Ontology and with species-specific phenotype ontologies such as those of the human, mouse, fly and worm. One of the main objectives of developing the Neuro Behavior Ontology is to discover the genetic basis of disease (Gkoutos *et al*., 2012). In BDEI the use of ontology modeling in the study of behavior, including behavioral ecology, has been addressed in an ontology of male jumping-spider courtship behavior and an ontology of sea turtle nesting behavior (Midford, 2004). In the latter case the ontology was informed by an ethogram of sea turtle nesting behavior, which codifies the animal’s behavioral repertoire. Whereas the conceptual stance of this work is comparable to ours in its emphasis on events, our work differs in two respects. Firstly, we adopt the event-centric perspective specifically to represent behavior that forms part of interspecific ecological interactions. Secondly, we model only the necessary knowledge of behavior that is required to create an ontology framework in a semantic enrichment and mediation system that integrates and transforms heterogeneous data into information. Other than the work mentioned above, our reading of the literature found no detailed work that focused on flower-visiting behavior or behavioral ecology or ecological interactions, neither in general nor of a specific group of species. There is thus great potential to extend the coverage of biodiversity and ecological concepts even further than the scope of the DSW, BCO and PCO ontologies described above, into the area of intersection between animal behavior and ecology (or behavioral ecology), and specifically into the domain of interspecific ecological interactions. What is needed is a conceptual model of behavior, behavioral ecology and ecological interactions that can be re-used easily, specifically by linking its classes to occurrence-records or specimen-records in typical biodiversity data-stores.

1. **Background to the case-study: biodiversity data quality in South African museums**

South African natural history museums participated in a program (Coetzer *et al*., 2012) to cleanse and migrate their data to a standard relational database schema and application (Specify Collections Management Software, University of Kansas Biodiversity Institute). Despite having general biodiversity data of a higher quality after the program’s conclusion, as well as syntactic interoperability, participating researchers of flower-visiting ecology were still unable to easily extract meaningful summaries across data-stores because semantic heterogeneity remained unresolved. The research reported here was therefore undertaken to integrate three selected data-stores containing data related to collections of flower-visiting insects, namely those of the Albany Museum (AM) in Grahamstown, Iziko South African Museum (SAM) in Cape Town and Plant Protection Research Institute (SANC) in Pretoria.

The Specify database schema is a powerful tool for expressing the structure and complexity of biodiversity data, particularly data on ecological interactions: the collection relationship table allows a collection object (specimen-record) in one collection to be related to one or more collection objects in the same or a different collection. Initially all three data-stores had only an arthropod collection, the collection objects of which included a field that may or may not have contained the species name of the plant with which the arthropod specimen was associated. The plant names were extracted from the arthropod collection objects and became collection objects in a new collection of plant observation records (whereas physical arthropod specimens are curated in collections by these museums, plant specimens are not). We established the same collection relationship, namely ‘host-plant’, between the arthropod specimen collection and the plant observation collection in each data-store. This allowed us to consistently represent the relationship between an arthropod herbivore specimen and the host-plant with which it was associated. Only arthropod records that had an associated host-plant record were processed further.

Table 1 summarizes the data attributes that characterized the standardized data-stores and shows how the word ‘flower(s)’ could be used to distinguish flower-visiting records. The heterogeneity of biodiversity information is evident in Table 1. For example, AM is a specialized flower-visiting data-store because it includes even the colours of visited flowers, and almost all the records are marked with the words ‘visit’ and ‘flower’. On the other hand, because there are very few records that have values in the [Behavior] field, SANC and SAM mostly contain information that is not as meaningful as the information in AM, though it is still useful.

**Table 1.** Data attributes from the three data-stores. FV = percentage explicit flower-visiting records. Flower-visiting records were distinguished by the [Sampling Method], [Behavior] and [Plant Part] fields.

|  |  |  |  |
| --- | --- | --- | --- |
|  | SAM data  3% FV(n=2 094) | SANC data  4% FV(n=219) | AM data  97% FV(n=21 159) |
| Host Type | host-plant | host-plant | host-plant |
| Host-plant | Diascia  capensis | Ruschia  indecora | Indigofera  nigromontana |
| Sampling Method | **Flowers** | swept from  **flowering** Acacia  albida | hand net |
| Behavior | Foraging  on nectar | [no data] | visiting  **flowers** |
| Plant Part | Leaf | **Flower** |  |
| Flower Colour | [no data] | [no data] | deep pink |

1. **Ontology development**

Ontology construction was informed by interviews with flower-visiting ecologists, who articulated the most important concepts, which were broken down into more specific concepts when necessary. Concepts were also created by reading relevant literature (top-down approach) and by examining flower-visiting data (bottom-up approach). Modeling in OWL was executed using the Protégé tool (Horridge, 2011) and in accordance with the middle-out ontology construction approach (Uschold & Gruninger, 1996).

* 1. Ontology construction in the domain of flower-visiting behavioral ecology

We limited our modeling to angiosperms (flowering plants) that are pollinated by vectors and not by an abiotic medium such as wind or water. We circumscribed as flower-visitors those taxa that belong to the phylum Arthropoda i.e. including the terrestrial groups represented broadly by spiders, millipedes (which mostly inhabit the soil) and insects. Plant galls caused by developing insect larvae, including larvae developing in flower-galls, were excluded from the domain, but the behavior of the adult insects which gave rise to these larvae was included in the domain.

Various kinds of animals, including arthropods (e.g. insects), birds (e.g. hummingbirds and sunbirds) and mammals (e.g. bats) are well-known *flower-visitors* because they live a life of actively, frequently and consistently seeking out flowers in order to utilize the flowers themselves or their products. The term ‘anthophilous’ denotes organisms that are often found on flowers for some reason, including to ambush prey (e.g. spiders). The most important flower products are nectar, pollen and oil, which are ingested or collected by flower-visitors. Insects are important flower-visitors and many insect groups have co-evolved as pollinators of plants.

For the purpose of ontology construction our definition of a flower-visitor was based on a review of flower-visiting insects (Kevan & Baker, 1983). Flower-visitors include arthropods that hide in flowers (e.g. thrips), camouflage themselves against flowers in order to ambush prey (e.g. mantids) or lay eggs in flowers (e.g. fruit flies). Referring to beetles, for example, Kevan and Baker (Kevan & Baker, 1983) state that ‘the predatory Adephaga are not flower visitors but, among the Polyphaga, notable flower visitors are Elateridae, Scarabeidae, Cleridae, Nitidulidae, Chrysomelidae, Staphylinidae, Meloidae, and Cerambycidae’. An insect can be a flower-visitor even if it does not ingest or collect nectar, pollen, oil (with or without terpene fragrance), resin, gum, anthers, ovules, seeds, petals or some other part of the flower or the entire flower.

Pollination is defined with varying granularity. A simple definition reads: ‘The transfer of pollen from an anther to a stigma’ (Raven *et al*., 1986). Some definitions emphasize that all pollination is ultimately an event (one-step process) because it consists of the act by which pollen is deposited on the pollen-receptive surfaces of a flower (or other reproductive structure such as a cone). In the typical case, pollination (cross-pollination) is a two-step process whereby a vector (‘carrier’) transfers pollen from the anther of one flower to the stigma of another flower (Raven *et al*., 1986). This is the definition that forms the basis of our conceptual model, though we do not model pollination as a simple, discrete event. We consider pollination to be a broader and more complex process that starts with the flower-visitor and its visit to a flower.

In the study of arthropod flower-visiting behavioral ecology, pollination may or may not be confirmed in a field setting. Confirmation of pollination requires closely following the flower-visitor and recording its behavior to see whether it actually transfers pollen onto the stigma. Thus, when ecologists refer to ‘pollination’ or a ‘pollinator’, unless otherwise stated, the word is usually used loosely to mean ‘inferred pollination’ or ‘potential pollinator’ or ’pollen vector’ (an organism that carries or transports pollen). Flower-visiting records are therefore the basic currency of ecologists who study flower-visiting and pollination because flower-visiting is easier to observe with high confidence.

It is generally accepted that pollen-transfer, both from the anther to a flower-visitor and from the flower-visitor to the stigma, is an accidental process (except in fig-wasps, which seem to undertake an intentional pollination ritual). A flower-visitor can become more-or-less covered in pollen, which it may then groom off the surfaces of its body using its tarsi (feet) and mouthparts, and pack into the scopa (hairy patch) on the hind leg, or store on the abdomen or in the crop. The pollen is then taken back to the nest and fed to the young (e.g. social bees) or deposited as nest provision for future young (e.g. solitary bees). Some plants, e.g. orchids and milkweeds, produce a pollinium (plural pollinia), or pollen-mass, borne on a sticky stalk that adheres to the flower-visitor’s body. The whole complex including the pollinium and stalk is called a pollinarium (plural pollinaria).

* + 1. Expert knowledge and implicit knowledge of behavioral ecology

Researchers of flower-visiting and pollination know implicitly that e.g. an adult beetle or fly or wasp of a certain taxonomic group (e.g. monkey beetles of the tribe Hopliini), or any bee (superfamily Apoidea) has only one reason to be associated with a plant, and that is to visit the plant’s flowers, usually to ingest or collect nectar or pollen or other flower products. Kevan and Baker (Kevan & Baker, 1983) listed known flower-visiting groups and we consider this knowledge to be typical expert knowledge (e.g. requiring knowledge of morphology and insect identification) that is generally accepted by virtue of being published in the literature.

The importance of implicit knowledge is even more pronounced in the particular case of bees of the genus *Rediviva*, consisting of 26 species that are endemic to South Africa, Lesotho and Swaziland. Female *Rediviva* bees collect oil from a small number of plant species (about 140 species in 14 genera) whose flowers produce oil to attract the female *Rediviva* bees in particular; or female *Rediviva* bees will ingest nectar from the flowers of any number of other plant species that produce nectar instead of oil (Pauw, 2006). The female bees collect and carry the oil using hairs on their especially-adapted, long front legs, and take the oil back to their nests as nest-provision (i.e. the egg is laid on the oil in the nest and the female that laid the egg then abandons the nest while the larva develops by feeding on the oil). Male *Rediviva* bees only visit flowers that produce nectar, which, like the females that visit ‘nectar plants’ (plant species that do not produce oil), they ingest to sustain themselves. A nectar-plant could be any flowering plant species, in the area that the bee frequents, that happens to have nectar in its flowers at the time. The words ‘visit’, ‘flower’ or ‘oil’ never occur among all the specimen-records in the SANC data-store that were created during the course of preparing two seminal articles on the famous *Rediviva* oil-collecting bees of southern Africa. On the other hand, only 6 out of 1664 SANC specimen-records do not include the sex of the bee. The reason for this is that pinned bees are small and so are their labels, and there is simply not enough space for unnecessary information. No information was lost within the museum, however, because an expert only needs to know the sex of the bee specimen and the plant species name (the key to knowing whether or not this is an oil-producing species) to know whether a *Rediviva* bee was seeking (or collecting) nectar or oil, and that it therefore must have been visiting flowers (Whitehead *et al*., 2008; Whitehead & Steiner, 2000) and potentially (and unwittingly) pollinating plants. The general subject of oil flowers and oil-collecting bees has been reviewed (Rasmussen & Olesen, 2000).

* + 1. Representing expert knowledge to simulate an expert

Our objective was to infer flower-visiting events from records of bee specimens using the information digitized from specimen labels as evidence. We also needed to use external, generally accepted and relevant knowledge that particular named groups of species (e.g. flies in the family Syrphidae) are known to be typical flower-visitors. Because it can be abbreviated or fragmentary, label information, while not external, may nevertheless need to be taken at face value as circumstantial evidence rather than absolute proof. In doing this we are doing nothing that a domain scientist would not do, and we therefore claim to make the same reasonable inferences that would usually be made by an expert who analyzes the data manually using her own knowledge.

## 4.2 Descriptions of three domain ontologies

In this section we describe the core flower-visiting (FV) domain ontology that we constructed for classes representing knowledge of flower-visiting behavioral ecology, as well as two smaller domain ontologies that we constructed, namely the known flower-visiting group ontology (KFG) and the *Rediviva* behavior ontology (RBH). Files containing these ontologies may be downloaded from http://africanpollination.org/ontology/

* + 1. The flower-visiting domain ontology (FV)

The richness of flower-visiting behavioral ecology knowledge is represented in a detailed subsumption hierarchy (Figure 1) that specializes the most generalized FV:*PlantAssociationEvent* class. An instance of this class is an event during which there is an assumed spatio-temporal association between an arthropod organism and a plant organism.

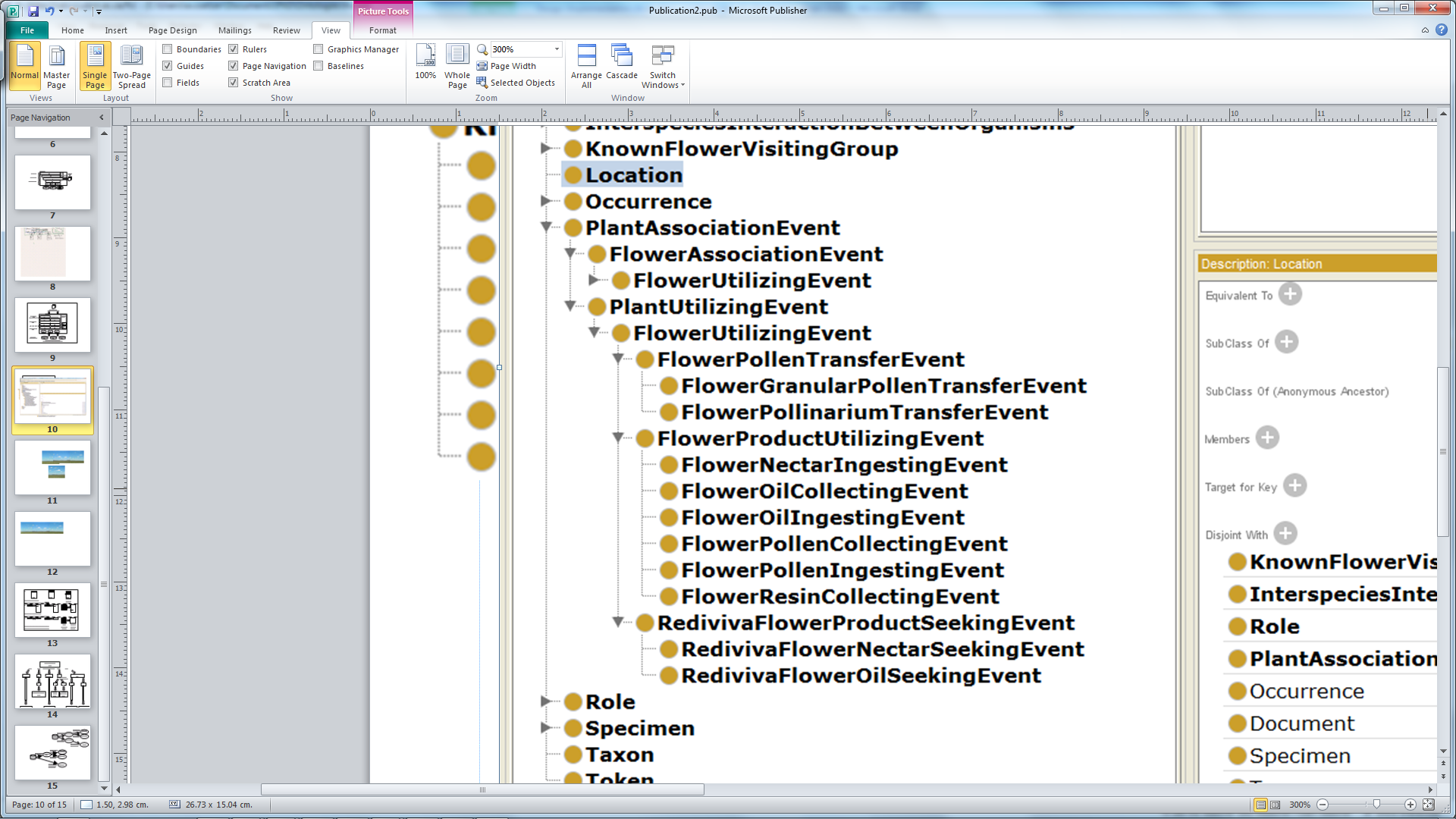
We adopt an event-centric perspective on animal behaviour and the ecological interactions signified by the behaviour. We therefore model different kinds of ecological events, including imprecise events such as associations between insects and plants (e.g. AssociationEvent), as well as more defined events elucidated by deeper interrogation of the available data and the application of expert knowledge (e.g. UtilizingEvent).

Central to this approach is the recognition that any concept is an event (e.g. a visit) rather than a physical object (e.g. an insect or a plant). Consider the following example. Suppose a class (a kind of event), *Event\_B*, is a subclass of another class, *Event\_A*. This means that an instance of *Event\_A* (the more general kind of event) will always occur when *Event\_B* (the more specific event) occurs. When a scientist observes and documents an insect sitting on a flower there are two conceptual events: the specific event of the insect sitting on the flower and the more generalized event of the insect sitting on the plant. The former event is not a part of the latter event and it does not come before or after the other event. Rather, the more specialized event *is*, at the same time, the more generalized event (which can nevertheless be conceived as an event in and of itself). We have more knowledge of the specialized event: Not only is the insect sitting on the plant but it is sitting in a specific region of the plant, i.e. the region occupied by the flower. The event-centric perspective allows us to see the relationships between the classes in Figure 1 (events) as subsumption, and allows us to detect and classify events from occurrences of specimens (things or physical objects).

The FV:*PlantAssociationEvent* class is specialized into the FV:*PlantUtilizingEvent* class, in which an instance is an event during which actual utilization of a plant was observed and recorded (e.g. by an expert’s description of the arthropod’s behavior as being ‘on plant’).

The FV:*FlowerAssociationEvent* class is a subclass of the FV:*PlantAssociationEvent* class. An instance of the FV:*FlowerAssociationEvent* class is an event during which there is an assumed spatio-temporal association between an arthropod and a flower.

**Figure 1.** The subsumption hierarchy representing detailed knowledge of flower-visiting behavioral ecology



Among the classes described thus far, a class name that contains the word ‘Association’ denotes the concept of a spatio-temporal association between an arthropod and a plant or flower, based only on the fact that a plant species name is included in the arthropod specimen-record. Our intention is to represent an assumed association between an arthropod and a plant because documented evidence is missing. On the other hand the presence of ‘Utilizing’ in a class name means that a plant or flower was observed being utilized by an arthropod, and that this was documented by the observer. Every instance of the FV:*FlowerAssociationEvent* class is also an instance of the FV:*PlantAssociationEvent* class. The reason for this is that a flower is a part of a plant, but, importantly, the object property is subsumption, and not *part\_of*. In other words, we model the event that occurs when the arthropod and the plant or flower are in contact (or are assumed to be associated), and not the detailed behavioral mechanism of the spatio-temporal relationship between the arthropod and the plant (e.g. ArthropodAppendage touches PlantSurface). While it is true that ecological specialization is the reason for many morphological modifications such as long legs or long mouthparts, such expert knowledge is not incorporated into the current event-centric conceptual model. These concepts are best modeled as continuants (physical objects) rather than occurrents (events in time), which opens up a future research avenue on the subject of how to reconcile these two perspectives in behavioural ecology.

The FV:*PlantUtilizingEvent* class is specialized into the FV:*FlowerUtilizingEvent* class, which is defined to contain instances of events, evidenced by direct human observations (recorded as notes), of the utilization by an arthropod of a flower surface (e.g. resting on a flower petal), flower space (e.g. ambushing prey inside a flower), flower tissue (e.g. chewing ovules) or flower product (e.g. ingesting nectar). Even a hovering moth that does not alight on the flower is utilizing the flower’s space by inserting its proboscis into the corolla tube. Multiple inheritance allows us to assert that the FV:*FlowerUtilizingEvent* class is a subclass of both the FV:*PlantUtilizingEvent* and the FV:*FlowerAssociationEvent*.

The more specialized subclass, FV:*FlowerProductUtilizingEvent*, subsumed by the FV:*FlowerUtilizingEvent* class, contains instances of events when the utilization of a flower product, such as nectar or pollen, was actually observed and recorded. The subclasses of the FV:*FlowerProductUtilizingEvent* class are therefore:

FV:*FlowerNectarIngestingEvent,* FV:*FlowerOilIngestingEvent,* FV:*FlowerPollenIngestingEvent,* FV:*FlowerOilCollectingEvent,* FV:*FlowerPollenCollectingEvent*, and FV:*FlowerResinCollectingEvent*.

Again, the event that occurs when an insect collects pollen from a flower is also, and will always be, an event that occurs when an insect sits on or inserts its mouthparts into (i.e. utilizes) a flower.

Because an instance of the FV:*FlowerPollenTransferEvent* class is passive or accidental, this class is not subsumed by the FV:*FlowerProductUtilizingEvent* class but by the FV:*FlowerUtilizingEvent* class. In other words, the flower support or flower space was actively utilized by the arthropod (e.g. the bee’s mouthparts penetrated the corolla tube) but the pollen was passively transferred. It would be incorrect to assert that the FV:*FlowerPollenTransferEvent* class is a subclass of the FV:*FlowerProductUtilizingEvent* class because this would mean that pollen can never be transferred without a flower product being utilized. The FV classes also reflect the fact that pollen may be granular or in a pollinarium, as described in Section 4.1. Since an arthropod may passively acquire both types of pollen, FV:*FlowerGranularPollenTransferEvent* and FV:*FlowerPollinariumTransferEvent* are not disjoint.

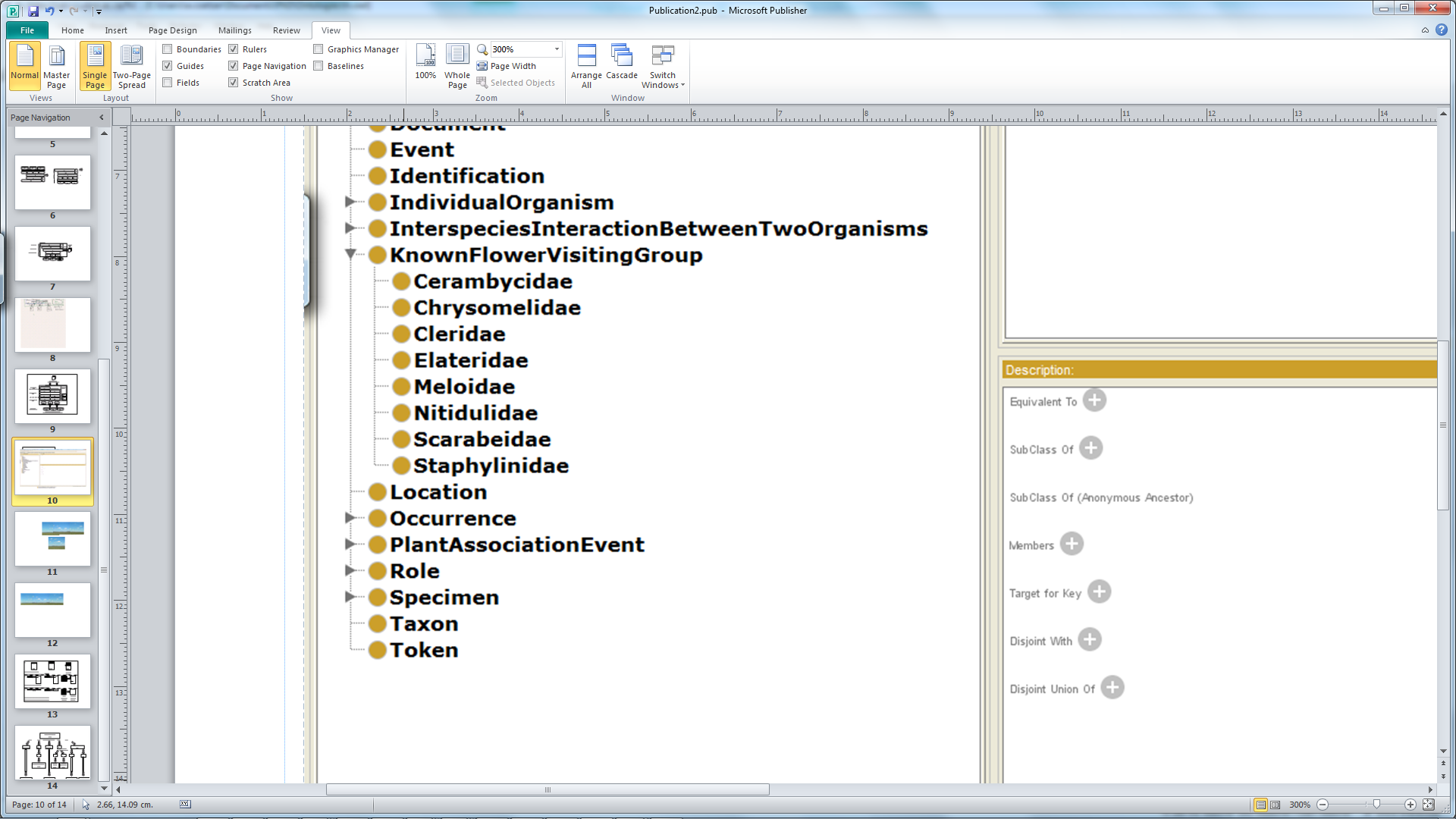
The FV:*FlowerUtilizingEvent* class is important with respect to the colloquial, domain concept of a ‘flower-visitor’. The event that occurs as a result of an observed and documented relationship between a flower-visitor and a flower (i.e. a putative ‘flower-visiting event’) is commensurable with the definition of the FV:*FlowerUtilizingEvent* class. At the same time the definition of a flower-visiting event (or that of a ‘flower-visitor’) may be broadened to be commensurable with the definition of the FV:*FlowerAssociationEvent* class. It may also be narrowed to be commensurable with the definition of the FV:*FlowerProductUtilizingEvent* class. Whereas domain scientists therefore commonly use one concept for a ‘flower-visitor’ or ‘flower-visiting event’, we use three concepts in a subsumption hierarchy for the event (Figure 1). This is discussed in the context of the evaluation of the system implementation, in Section 6.3 below. We further assert that a ‘plant-visiting event’ may similarly either be a FV:*PlantAssociationEvent* or a FV:*PlantUtilizingEvent*, depending on whether evidence has been documented.

The purpose of the subsumption hierarchy in Figure 1 is to instantiate the most specific event that can be justified with the evidence at hand. The most specialized events are the most important events because they characterize the ecological interactions.

* + 1. The known flower-visiting group domain ontology (KFG)

The KFG ontology (Figure 2) contains the KFG:*KnownFlowerVisitingGroup* class. Its subclasses are the names of groups of different ranks (e.g. family or tribe) consisting of species that are generally accepted to be typical flower-visitors as defined by Kevan and Baker (Kevan & Baker, 1983). The function of the KFG ontology is to enrich records from the data-stores by instantiating the *FV:FlowerAssociationEvent* class when an arthropod species is a member of a known flower-visiting group, and no other documented information indicates that flower-visiting took place. The assumption is that a plant species name would not be included in a data-store record of an arthropod specimen belonging to a group that is a known flower-visiting group (e.g. bees, in the family Apidae) if the arthropod specimen had not been ecologically associated with the flower of a specimen of the plant species.

**Figure 2.** A fragment of the KFG ontology representing generally accepted knowledge of the known flower-visiting groups of insects



* + 1. The *Rediviva*-behavior domain ontology (RBH)

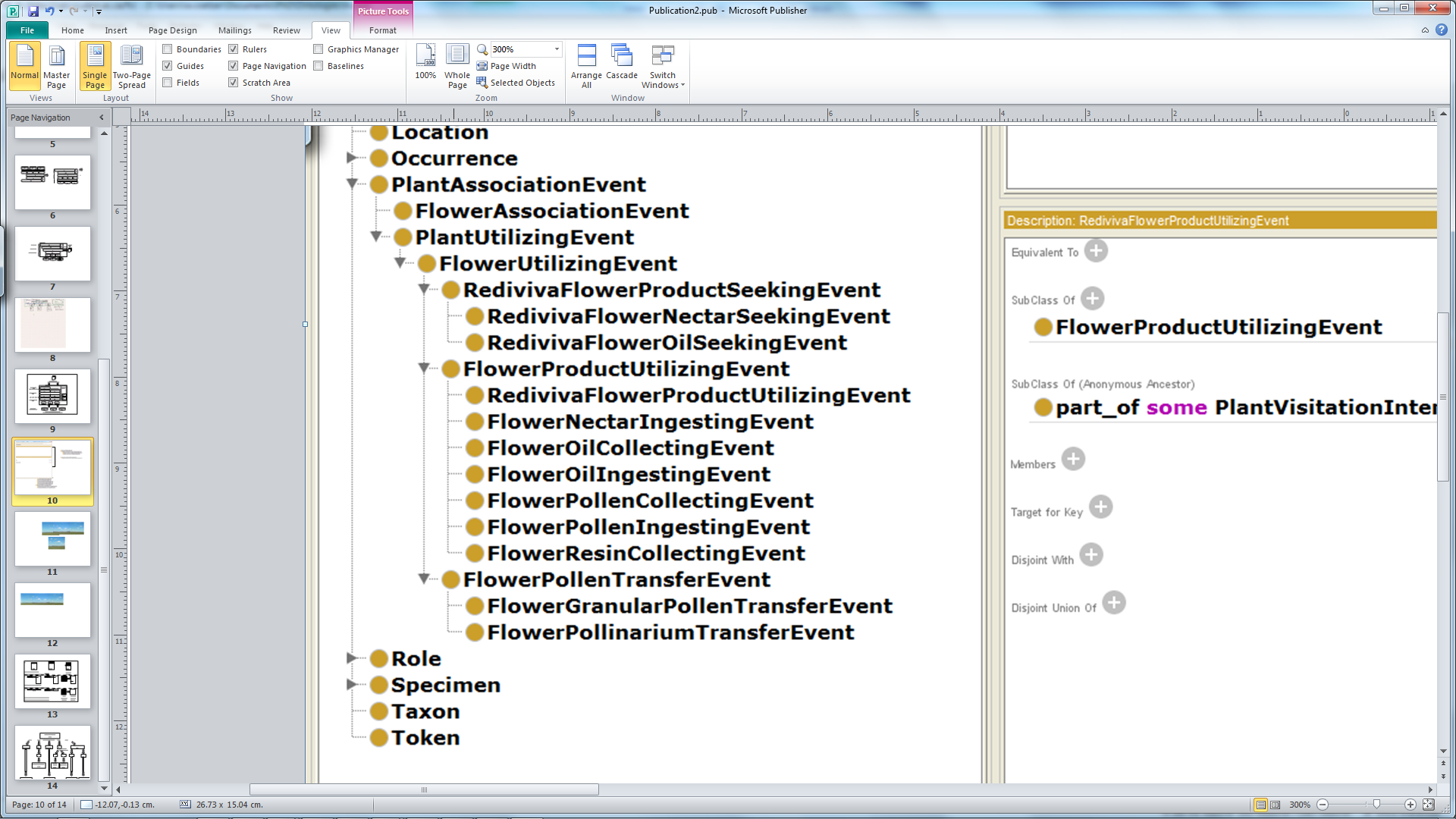
Within our scope, knowledge of the flower-visiting behavioral ecology of *Rediviva* bees can be summarized as follows. A plant species is either an oil-producing species or it is not an oil-producing species. Male and female *Rediviva* bees would not visit the flowers of plants not belonging to oil-producing species if they are not seeking nectar or ingesting nectar, and female *Rediviva* bees would not visit the flowers of plants belonging to oil-producing species if they are not seeking oil or collecting oil.

The RBH ontology (Figure 3) imports the FV:*FlowerUtilizingEvent* class, and specializes this class into the RBH:*RedivivaFlowerProductSeekingEvent* class, which subsumes the RBH:*RedivivaFlowerOilSeekingEvent* class and the RBH:*RedivivaFlowerNectarSeekingEvent* class.

A typical case is that of a female *Rediviva* bee observed alighting on the flower of a plant of an oil-producing species. We know that the bee is seeking floral oil even if we do not see the bee actually collecting or ingesting floral oil (i.e. utilizing a floral product). If the sex of the bee is unknown we still know that the bee is seeking a floral product.

Importantly, we can only assert, using the knowledge described above, that *Rediviva* bees seek floral products or that they seek oil or nectar, and not that any other arthropods seek these things. The RBH:*RedivivaFlowerProductSeekingEvent* class and its subclasses are useful because they allow us to enrich records of *Rediviva* bees to a specific class without observations detailing the behavior of a *Rediviva* bee collecting oil or ingesting nectar. We cannot do this with records of arthropods other than *Rediviva* bees. For this reason the FV ontology does not contain a class for representing the seeking of floral products by arthropods other than *Rediviva* bees.

**Figure 3.** The RBH ontology represents knowledge of the flower-visiting behavioral ecology of *Rediviva* bees



4.3 Linking expert knowledge to common biodiversity concepts

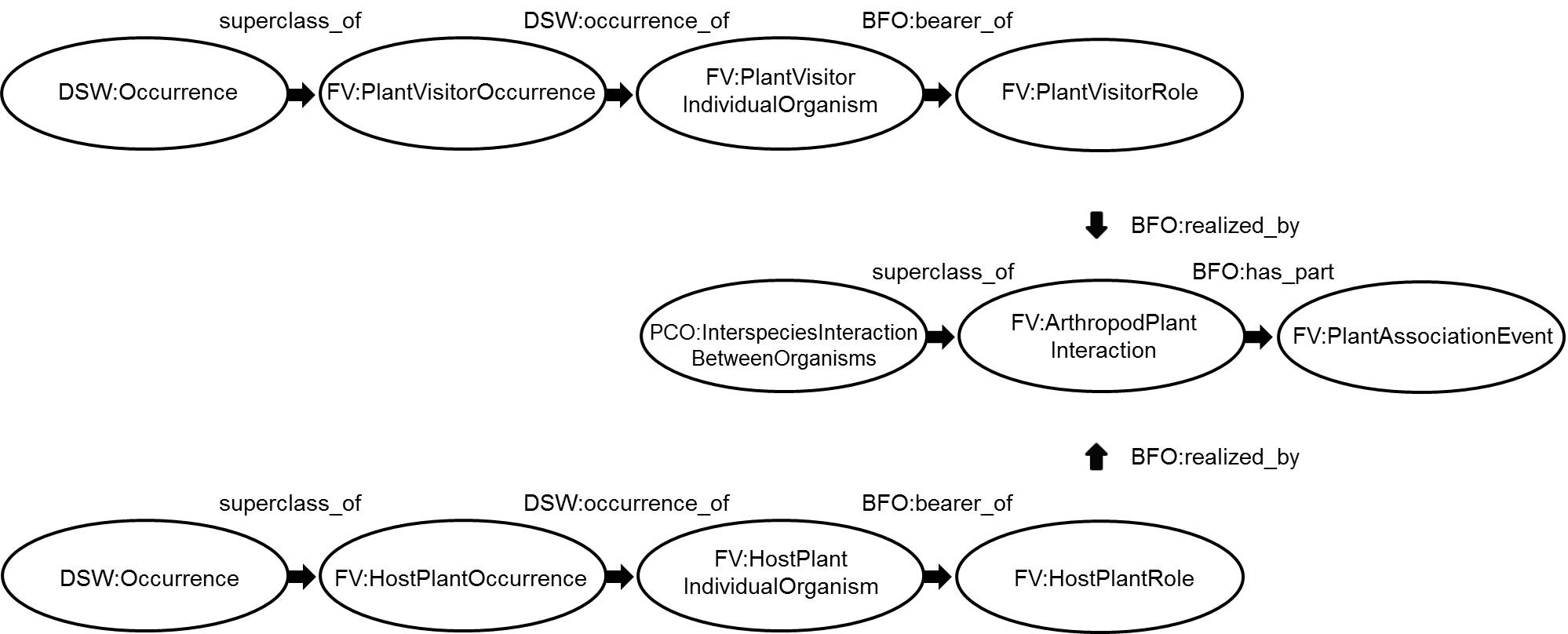
We used two external domain ontologies, namely Darwin-Semantic Web (DSW) (Baskauf & Webb, 2014) and the Population and Community Ontology (PCO), which complies with BFO. We used DSW to express the concepts (e.g. DSW:*IndividualOrganism* and DSW:*Occurrence*) commonly needed for rich semantics in the domain of specimen collections. The choice of the PCO ontology was important because we ultimately needed to express a flower-visiting event (e.g. an instance of the FV:*FlowerAssociationEvent* class) as a BFO:*part\_of* a PCO:*InterspeciesInteractionBetweenOrganisms* process.

The PCO:*InterspeciesInteractionBetweenOrganisms* class is a subclass of the BFO:*Process* class. We specialized the PCO:*InterspeciesInteractionBetweenOrganisms* class into the FV:*ArthropodPlantInteraction* class. An instance of the FV:*PlantAssociationEvent* class was asserted to be a BFO:*part\_of* an instance of the FV:*ArthropodPlantInteraction* class (Figure 4).

The Basic Formal Ontology (BFO) provided the *role* class (Arp & Smith, 2008), such that an independent continuant (e.g. an instance of the FV:*PlantVisitorIndividualOrganism* class) is the BFO:*bearer\_of* an instance of the FV:*PlantVisitorRole* class, which is BFO:*realized\_by* an instance of the FV: *ArthropodPlantInteraction* class. Similarly an instance of the FV:*HostPlantIndividualOrganism* class is the BFO:*bearer\_of* an instance of the FV:*HostPlantRole* class, which is BFO:*realized\_by* the same instance of the FV: *ArthropodPlantInteraction* class (Figure 4).

It was important to distinguish the (part of a) process (i.e. the FV:*FlowerAssociationEvent*) from the material entity (i.e. the individual arthropod organism) bearing the role (FV:*PlantVisitorRole* class) that realized the interaction process with the complementary material entity (i.e. the individual plant organism). This event-centric view on flower-visiting behavioral ecology, while having an intuitive scientific appeal, especially allowed the different kinds of arthropod-plant interactions, plant-visiting events and flower-visiting events to be extracted as the salient features.

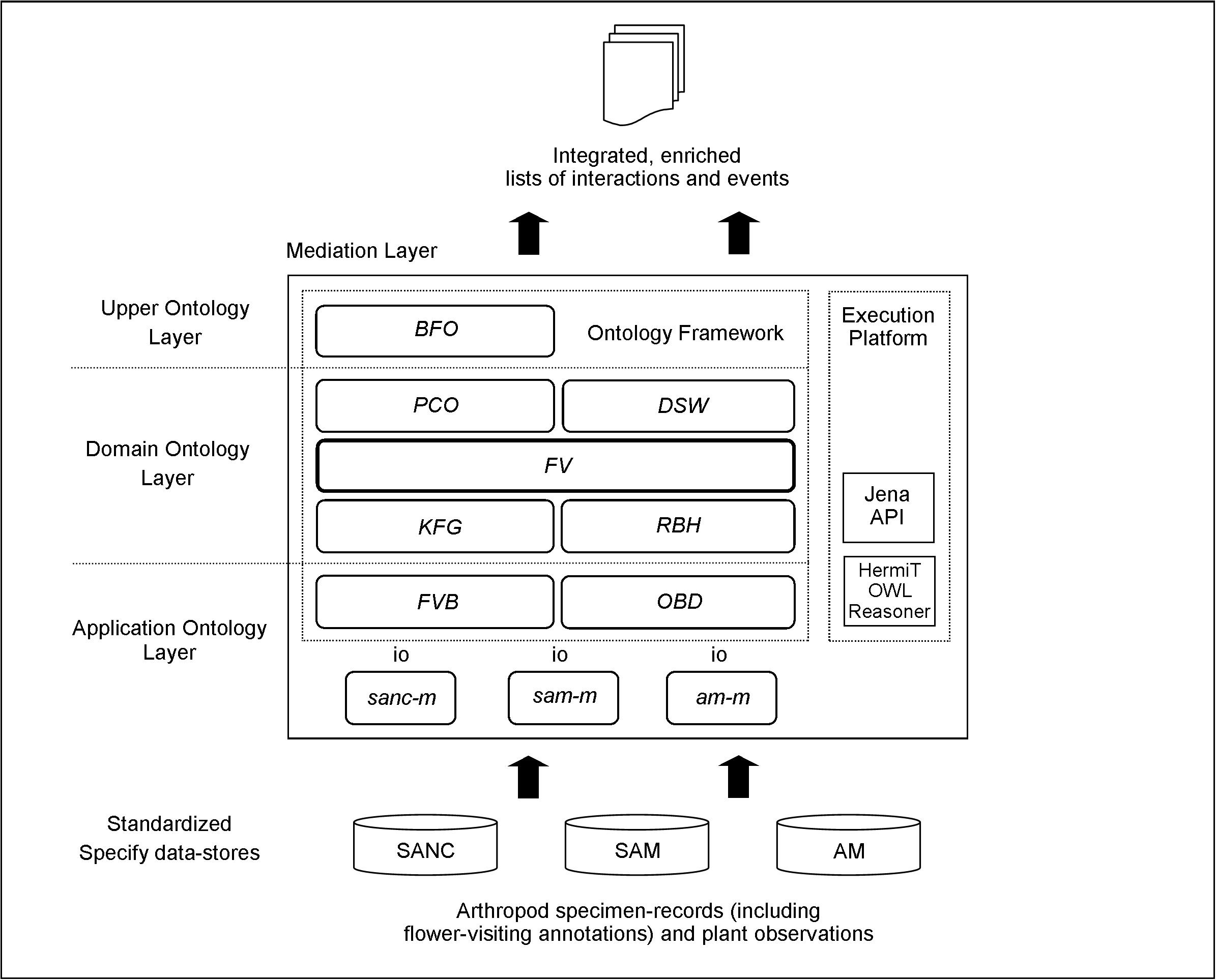
**Figure 4.** A fragment of the FV ontology showing object properties and classes giving rise to the FV:*PlantAssociationEvent* class. Common biodiversity concepts, above and below, are distinguished from the expert, and often implicit, ecological concepts between them, which are seen from a higher-level perspective. Abbreviations: BFO‒Basic Formal Ontology; DSW‒Darwin Semantic Web Ontology; FV‒Flower-visiting Ontology; PCO‒Population and Community Ontology



1. **Mediation system**

We designed and implemented a prototype system for semantic enrichment and mediation that uses the ontologies described above. The mediation system automates the transformation and integration of heterogeneous flower-visiting data into meaningful ecological information. The architecture of the mediation system is depicted in Figure 5. The mediation layer is responsible for integrating and transforming data from the three data-stores into standardized biodiversity information that is semantically enriched with ecological knowledge. The mediation layer includes the execution platform and the ontology framework, consisting of the application ontologies, domain ontologies and the upper ontology. Rather than being the final implementation, we consider this to be an early version that may be modified in future to allow for the inclusion of more detailed knowledge or even a different conceptual stance.

**Figure 5.** Architecture of the semantic enrichment and mediation system. Abbreviations ‒ BFO: Basic Formal Ontology; DSW: Darwin Semantic Web Ontology; FV: Flower-visiting Ontology; PCO: Population and Community Ontology; KFG: Known Flower-visiting Group Ontology; RBH: Rediviva Behavior Ontology; FVB: Flower-visiting Behavior Ontology; OBD: Observation Date Ontology



5.1 Application ontologies

The mappings and application ontologies link the data in the data-stores to classes in the core FV domain ontology. The Observation Date Ontology (OBD) is specific to the AM data-store. The FVB application ontology has a distinct mapping to each data-store (e.g. sanc-m).

* + 1. The flower-visiting behavior application ontology (FVB)

The function of the flower-visiting behavior application ontology (FVB) is to classify a record from a data-store to the most specialized subclass of the FV:*PlantAssociationEvent* class that is justified: the FV:*FlowerAssociationEvent* class in the case of more general behaviors or one of the latter’s subclasses in the case of more specific behaviors. The FVB ontology mapping (Table 2) therefore contains all the data-store fields that could potentially contain words indicating that flower-visiting had been observed, namely [Behavior], [Plant Part], [Sampling Method], and [Observer Name]. Classes in the FVB ontology include literal text strings originally written in field notebooks by observers and stored in the [Behavior] field, which describe the behavior of arthropods when visiting flowers.

The definitions of the [Plant Part] and [Sampling Method] fields are similar. The former means that a part of the plant (e.g. a flower) was the subject of the observation and the latter means that the method of the observation was to focus on a part of the plant. Values in the [Plant Part] and [Sampling Method] fields in the data-stores were used to create FVB classes only when the [Behavior] field had no value (i.e. if values were present in all three fields, or only in the [Behavior] field, only the [Behavior] field was used).

The rationale for using the [Observer Name] field (the values of which are classified as ‘expert’ or ‘non-expert’) is that, provided that the arthropod species belongs to a known flower-visiting group, the names of expert observers are good indicators of observations of behavior that correspond to the definition of the (more specific) FV:*FlowerProductUtilizingEvent* subclass, even if no other data are present.

**Table 2.** Partial lists of the FVB mappings from the three data-stores (displayed as a single table). The FVB mapping instantiates the FV:*PlantAssociationEvent* class or its subclasses on the basis of the [Behavior], [Plant Part], [Sampling Method] and [Observer Name] fields.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data-store | Plant Part | Sampling  Method | Observer  Name | Expert | [Behavior] field | FV Class |
| sam-m |  |  |  |  | Collecting pollen on yellow flowers | *FlowerPollenCollecting*  *Event* |
| sam -m |  |  |  |  | Feeding on Brunia laevis pollen | *FlowerPollenIngestingEvent* |
| sam -m |  |  |  |  | Foraging on nectar of Euphorbia flowers | *FlowerNectarIngestingEvent* |
| sam -m |  |  |  |  | Taking resin from Dalechampia capensis | *FlowerResinCollectingEvent* |
| sam -m |  |  |  |  | Visiting extra-floral nectaries | *PlantUtilizingEvent* |
| am-m |  |  |  |  | On foliage | *PlantUtilizingEvent* |
| am -m |  |  |  |  | On stem of plant | *PlantUtilizingEvent* |
| am -m |  |  | Fred  Gess | Yes | Visiting flowers | *FlowerProductUtilizingEvent* |
| am -m |  |  | Fred  Gess | Yes |  | *FlowerProductUtilizingEvent* |
| am -m |  |  | Vernon  Smith |  | Visiting flowers | *FlowerUtilizingEvent* |
| am -m |  |  |  |  | In flowers | *FlowerUtilizingEvent* |
| am -m |  |  |  |  | On flowers | *FlowerUtilizingEvent* |
| am -m | Flower |  |  |  |  | *FlowerAssociationEvent* |
| sanc-m |  | Flowers |  |  |  | *FlowerAssociationEvent* |

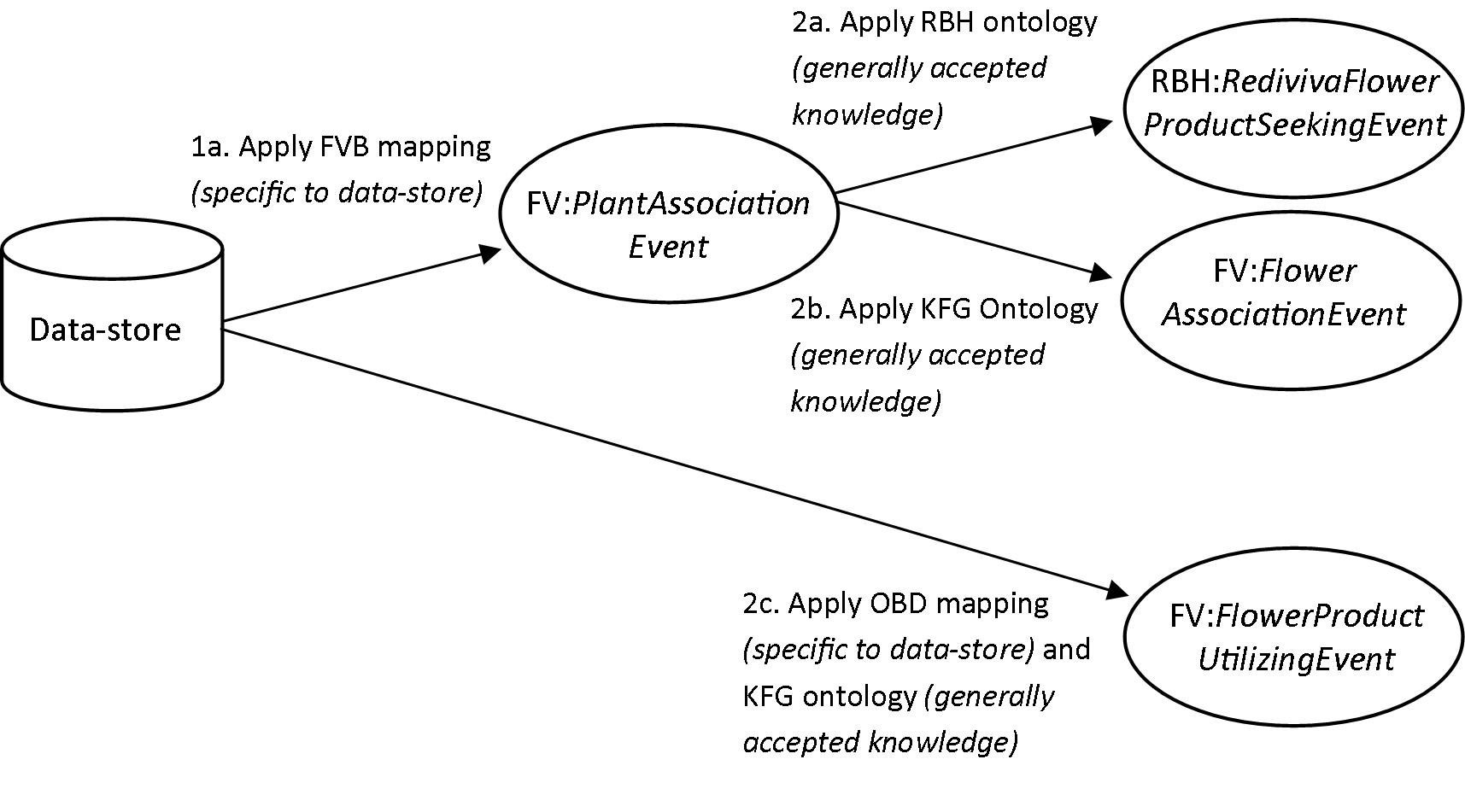
* + 1. The observation-date application ontology (OBD)

The observation-date application ontology (OBD) is specific to the AM data-store. The records in the OBD ontology are ranges of observation dates. The rationale for using this ontology is that the AM data-store is so specialized that mere enrichment to the FV:*FlowerAssociationEvent* class would be too broad in the context of what is known about the detailed, expert information contained in the AM data-store. The reason for this is that these were times when a handful of known experts were active, and we know that they followed a particular field sampling routine, namely collecting insects that were collecting or ingesting pollen or nectar from flowers, even if they did not record this behavioral information.

* 1. The execution platform

A prototype execution platform was implemented. This used the Java Jena API to instantiate the ontology classes. This occurred as a result of a string comparison between the value read from the data-store and a class in the FVB or OBD application ontologies (via the respective mappings). Instantiation occurred in two steps, first using information specific to the data-store (Figure 6, step 1a or 1b) and then using generally accepted knowledge (Figure 6, step 2a or 2b or 2c). Step 1b occurred in the case of AM data-store records of species of known flower-visiting groups that had an observation date but no observer name (In step 1b the FV:*PlantAssociationEvent* class is always instantiated). In Figure 6 the class that is instantiated is the most generalized class that can be instantiated using the mapping (FVB mapping) or ontology (FV, KFG or RBH ontology) that is shown. For example, after step 1a more-detailed behavioral information may result in the instantiation of the FV:*FlowerPollenCollectingEvent* class.

**Figure 6.** Instantiation occurs in two steps, first using information that is specific to the data-store, in an application ontology, and then using generally accepted knowledge in a domain ontology



* 1. System output: linking the domain ontologies to obtain integrated ecological information

When the execution platform instantiates the FV:*PlantAssociationEvent* class it also instantiates the FV:*PlantVisitorOccurrence* and FV:*HostPlantOccurrence* classes. The class linkage is completed by instances of the FV:*PlantVisitingIndividualOrganism*, FV:*PlantVisitorRole* and FV:*ArthropodPlantInteraction classes*, as well as instances of the FV:*HostPlantIndividualOrganism* and FV:*HostPlantRole* classes (Figure 4).

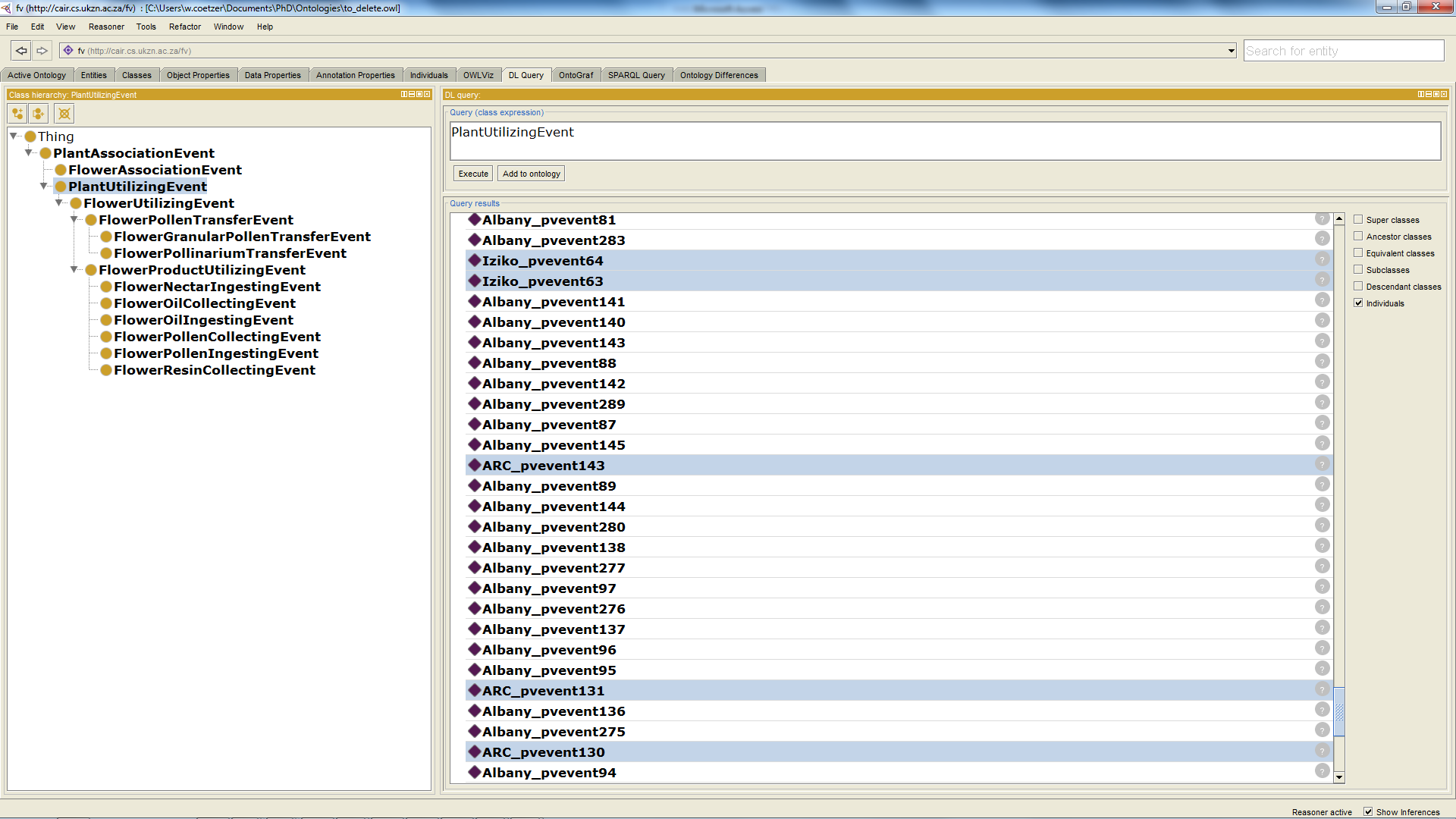
Invoking the HermiT reasoner results in the classification of instances of the subclasses of the FV:*PlantAssociationEvent* class, which yields integrated lists of instances, from all three data-stores, of the different subclasses of the FV:*PlantAssociationEvent* class. These are constituted into arthropod-plant interactions. For example, a researcher may export a list of instances, from all three data-stores, of the FV:*FlowerNectarIngestingEvent* class, or ‘times when arthropods were observed ingesting floral nectar’. Because the ontology represents the fact that an arthropod ingesting floral nectar must be visiting a flower (and therefore a plant), no further manipulation is required to include these instances in more general lists of times when a plant visitor (i.e. an instance of the FV:*PlantVisitorIndividualOrganism* class) was observed:

‘utilizing a floral product’, or ‘utilizing a flower’, or ‘utilizing a plant’ (Fig. 7), or, more generally, times when an arthropod was:

‘associated with a flower’, or ‘associated with a plant’, or, at the most generalized level, times when:

an arthropod and a plant were involved in an ecological interaction.

**Figure 7.** The results of a Description Logics query of the FV (flower-visiting) knowledgebase using the Protégé application. The query requests a list of instances of the FV:*PlantUtilizingEvent* class among all three data-stores. Instances from the Iziko and ARC data-stores are highlighted in the integrated list, which includes mostly Albany data-store instances.



1. **System evaluation and discussion**

For the purpose of automating data integration the mediation system needed to:

1. Capture the essential data elements and background knowledge, e.g. the general concept of a biodiversity specimen-record, or instance of the DSW:*Occurrence* class;
2. Make explicit the specific knowledge that an expert would manually extract from the data, e.g. that the organism that became preserved as the specimen was observed behaving in a particular way which meant that it was interacting with another organism, and that this interaction was of a particular type;
3. Transform the data, which emphasize arthropod specimens and plant observations, into events and interactions, which emphasize the ecological relationships between arthropods and plants.

We have demonstrated that semantic enrichment and mediation, as implemented in the described system, can be used to automate the integration and transformation of records of flower-visiting behavioral ecology among arthropod specimen-records from different data-stores, in which the flower-visiting information, specifically, is implicit, fragmented or heterogeneous, or even missing. The system output is integrated lists of enriched plant-arthropod interactions and the plant- and flower-visiting events constituting these interactions.

In the AM data-store the information about flower-visiting behavior was mostly in the [Behavior] field, which contained text strings describing arthropod behavior. In the SAM and SANC data-stores there were few values in the [Behavior] field and the [Sampling Method] and [Plant Part] fields also contained flower-visiting information, though this needed a degree of semantic enrichment in which we did not have high confidence. In this section we explore the potential for, and limitations of, semantic enrichment and mediation. We evaluate the application ontologies that link the data to classes in the FV ontology, and we consider the resolution of missing data and the extraction of new information from the data. We also reflect on how automated integration, as performed by the mediation system, may compare to manual integration by one or more experts. We discuss the scalability, extension and potential impact of the mediation system.

* 1. The potential for, and limitations of, semantic enrichment and mediation

The potential for semantic enrichment and mediation differed between the data-stores. The AM data-store had the most potential for enrichment to a specific class (namely FV:*FlowerProductUtilizingEvent*) in most records in the data-store. The reason for this was that the specimen collection itself and the specimen-records had been assembled and written down by a relatively small number of people, mostly experts, who also described the historical development of the collection. The objectives and methods of these experts, in building the specimen collection and recording the data, were focused on flower-visiting behavioral ecology, and consistent, and the data were of a high quality in the sense that the records contained meaningful behavioral information. On the other hand the SAM and SANC data-stores needed semantic enrichment in most records even to instantiate the a general class, namely the FV:*FlowerAssociationEvent* class, because of missing data. These data-stores are associated with specimen collections that were built by much larger and more diverse groups of people who had more diverse research objectives and sampling protocols. The *Rediviva* bee specimens in the SANC data-store were an exception to the general pattern of missing data in the SANC data-store because the potential for semantic enrichment of these records (though they were small in number) was probably the highest of all. The reason for this was that these records documented specimens of species that exhibit very strong evolutionary relationships with oil-producing plants, and this is well understood and recorded in the scientific literature.

The following cases of semantic enrichment were designed and implemented, and we now consider their inherent assumptions. The three cases differ in that the need for semantic enrichment had a different origin in each.

1. Known flower-visiting group: If the arthropod species to which a specimen belonged was a member of a known flower-visiting group (e.g. family), and only the associated plant species name was given (i.e. there was no other recorded label information about the relationship of the specimen to any part of the plant, or its behavior), then the record was enriched to be that of a FV:*FlowerAssociationEvent*. Since this assumption was made because of missing data, our confidence in making the assumption was lower than our confidence in the semantic enrichment discussed in 2 and 3 below.

Nevertheless, there were grounds for justifying this assumption. Flower-visitors need to visit flowers frequently in order to ingest pollen or nectar to sustain themselves, or collect flower products with which to provision their nests or feed their young, or have a platform for prey-ambush. Moreover, many flower-visiting ecologists do not record the behavior of arthropods because it is assumed that posterity will accept implicitly that their specimens were flower-visitors because they were known to be flower-visiting ecologists. The instances of the FV:*FlowerAssociationEvent* class were therefore inferred. This is a relatively general class, being subsumed directly by FV:*PlantAssociationEvent*. The latter class is the most generalized class in the subsumption hierarchy, and was instantiated in every dataset row.

1. Flower product utilization: We understood the criteria necessary (i.e. which expert observers or which range of observation dates, in the absence of observer names) to instantiate the specific class of FV:*FlowerProductUtilizingEvent* instead of the more general FV:*FlowerUtilizingEvent* or FV:*FlowerAssociationEvent* (as long as the arthropod species belonged to a known flower-visiting group). The semantic enrichment function discussed below was separated into two application ontologies, namely FVB and OBD, because the use of the observation date was restricted to the AM data-store. Our confidence in these assumptions was very high, as explained below.

The [Observer name] field: We inferred that the expert observed a specific behavior, indicating that the arthropod had been utilizing (ingesting or collecting) a flower product (nectar, pollen or oil) even if the behavior had been originally recorded using the more general phrase ‘visiting flowers’. This phrase is a common way of abbreviating a complex behavioral observation in the Albany data-store. This assumption was justified because particular expert observers have accumulated the necessary expertise, over an entire career, to recognize the specific behavior that constitutes the utilization of flower products. After all, ten years ago an observer may not have realized the importance of recording the specific behavior that was observed (e.g. ‘spending time on flower ingesting either nectar or pollen’), even if time or convenience allowed this. A record annotated with the phrase ‘visiting flowers’ by an inexperienced observer, however, will cause a more general instance of the FV:*FlowerUtilizingEvent* class to be instantiated. This class is still more specific than would result from enriching an inexperienced observer’s record with no flower-visiting information but where the arthropod species belongs to a known flower-visiting group (i.e. FV:*FlowerAssociationEvent*).

The [Observation date] field: We know the history of the specimen collection and that records collected during particular ranges of dates originated from expert observers who were collecting insects that were ‘visiting flowers’ (meaning that the insects were utilizing flower products), even if these observers’ names were not recorded on the labels or in the database records. In contrast the same confidence cannot be attributed to records collected at other times, or more recently (i.e. by relatively inexperienced observers) and these would therefore need to be more detailed if they are to be interpreted to mean that a more specific behavior had been observed.

1. Oil seeking behavior versus nectar seeking behavior by *Rediviva* Bees: The case of female and male bees in the genus *Rediviva* is a typical case of expert knowledge in natural history, where behavioral ecology has generated predictive knowledge on the basis of strong evolutionary relationships between interacting species. As described in Section 3.4.3, a record was therefore enriched to be that of a FV:*FlowerOilSeekingEvent* only if the plant belonged to a known oil-producing species and the bee specimen was a female, and that of a FV:*FlowerNectarSeekingEvent* if the bee specimen was a male (irrespective of whether or not the plant belonged to a species that produces oil), or if the bee specimen was a female and the plant species is known to not be an oil-producing plant species.

Our confidence in this implicit knowledge and these assumptions was very high. This reflected the general acceptance of this knowledge in the scientific community. In fact, whether or not a bee was collecting nectar or oil was never recorded on the museum specimen labels or elsewhere because this was unnecessary. For this reason these data (whether oil or nectar was being sought or collected) were not considered to be missing.

The *Rediviva* behavioral ecology information was new information that was not explicit in the data. This new information was extracted from the data through the representation, in the RBH ontology, of expert knowledge of the behavioral ecology of *Rediviva* bees.

* 1. Manual mapping of the flower-visiting behavior ontology (FVB)

Compared to the RBH and KFG domain ontologies, the FVB and OBD application ontologies were less re-usable because the knowledge represented by classes in these ontologies was not generally accepted knowledge. The complexity of arthropod behavior, and flower-visiting behavior in particular, required classes in the FVB ontology to be manually mapped to classes in the FV ontology. This was done by translating literal text strings describing arthropod behavior (recorded in the data-stores) into FV ontology classes. Our scope did not allow us to develop a complete model of the behavior of arthropods visiting flowers, and neither is such a complete view of behavior available or perhaps even possible. Rather our position was that a circumscribed body of arthropod flower-visiting behavioral observations, partly describing flower-visiting interactions, was available, and that these partial observations could be integrated, at least partly by automated means.

The implementation decision to map FVB classes to FV classes by manual means will result in the need for the FVB ontology to be edited manually whenever a new behavior value is added to a data-store. The new behavior class will also need to be mapped manually to an FV ontology class by a flower-visiting expert. The advantage of this design is that expert input will continue to enrich the FVB ontology with classes that are as specialized as is possible. It also means that the engineering and development process will remain more adaptable to the needs of experts and users rather than becoming constrained and thereby ultimately limiting the system’s utility.

* 1. Resolving the problem of missing data

Missing data forced us to find more or different evidence of flower-visiting by mapping additional data-store fields to classes in the FVB application ontology. Our confidence in our assumptions was very high in the case of enrichment in *Rediviva* behavioral ecology data (RBH ontology) and that of expert knowledge of flower product utilization (the [Observer Name] field in the FVB ontology, and the OBD ontology). Neither of these were cases of missing data, but rather cases where data could be confidently supplemented with knowledge or enriched.

We had low confidence in the case of missing behavioral data which forced us to infer that a flower-visiting event had occurred if a species belonged to a known flower-visiting group (KFG:*KnownFlowerVisitingGroup* class). Also, if no data were present in the [Behavior] field it was necessary to query the [Plant Part] and [Sampling Method] fields. By querying these fields, however, we may have inadvertently introduced records of immature arthropods developing in flowers rather than the sought-after adult arthropods visiting flowers. When entomologists survey the fauna of plants they often collect flowers and allow the immature arthropods developing in the flowers to complete their life-cycles and emerge as adults. In such a case the word ‘Flower’ would appear in the [Plant Part] field. One could try to limit the assumptions by adding yet another field to the FVB ontology, namely [Life-Stage], and exclude records that contained words such as ‘larva’ or ‘caterpillar’. Soon, however, a pattern of ‘missing data begets missing data’ is established, which is ultimately caused by a lack of rigour in the original field sampling, annotation, database design or data capture. It was for this reason that the FV subsumption hierarchy contained a specialized subclass of the FV:*PlantUtilizingEvent* class, namely the FV:*FlowerUtilizingEvent* class, which was instantiated only through directly observed flower-visiting behavior (whereas the FV:*FlowerAssociationEvent* class was instantiated through semantic enrichment of records of arthropods belonging to known flower-visiting groups or through the FVB ontology when the word ‘flower’ appeared in the [Plant Part] or [Sampling Method] fields). A user can therefore limit an analysis to include only instances of the FV:*FlowerUtilizingEvent* class, and thereby avoid any assumptions that were made to remedy missing data, though at the cost of analyzing fewer data from fewer data-stores.

* 1. Comparison between automated and manual data integration

A relatively high level of automation was achieved, but the design of the FVB application ontology increased the overhead cost by requiring manual representation and mapping of new text strings describing arthropod behavior. On the other hand, the FVB ontology, which represents arthropod behavior, is the key to high-quality semantic enrichment and mediation within the defined scope. On balance the manual input into the FVB ontology is seen as a strength rather than a weakness because of the inevitable need to reduce the complexity of representing arthropod behavior through the input of a scientist.

While no empirical comparison between manual and automated integration was conducted, we believe that automated data integration as performed by the system will:

1. be more objective and consistent than a manual integration effort, especially where this is undertaken manually by more than one person;
2. include more expert knowledge than would be included in a manual integration by a scientist with a lower level of expertise because an expert creates and edits the FVB ontology;
3. allow the user to exclude assumptions borne of missing data, whereas this may not be true of a manual integration project.
   1. Scalability and extension

To the extent that the FVB application ontology contains manually created classes and uses manually created mappings the implemented system is a specific solution for the three particular data-stores that were used in this study. There was, however, little variability in overall structure between the data-stores, which were typical specimen databases from traditional natural history museums. Such databases contain fields that mostly represent the provenance (e.g. date or locality) and biological classification of stored specimens for the purpose of basic collection management (e.g. inventory and curation) and as evidence to use when describing, naming and classifying species (taxonomy and systematics). This reflects a tradition of natural history museums that is about 260 years old. Finding detailed behavioral or interaction information in biodiversity databases is the exception rather than the rule, and it is only recently that models (such as the Specify database schema) for representing richer, deeper and more extensive biodiversity or ecological information, including biotic interactions, have been developed and adopted.

The variability of flower-visiting data-stores is likely to be more constrained than that of biodiversity data-stores in general. It could therefore be relatively easy to add a fourth data-store to the system or to use the system to integrate data from many distributed flower-visiting data-stores. For this reason we expect the described system implementation to be widely applicable in studies of flower-visiting behavioral ecology. The specific task of engineering interoperability among distributed data-stores will be the subject of future research.

* 1. Potential impact

The event-centric approach to ontology construction could be important for semantic interoperability in behavioral ecology, especially in cases where the complexity of representing knowledge of animal behavior and ecological interactions needs to be reduced or abstracted, and where large datasets of specimen-records need to be integrated. In the field of environmental science Villa *et al*. (2009) showed how declarative modeling can incorporate a knowledge model to produce a ‘semantically aware environmental model’, which suggests that the approach described above may be useful in such an application.

Analyses of vertebrate or invertebrate stomach contents or relationships between occurrences of intertidal invertebrates may lend themselves to this event-centric approach to conceptual modeling. Among arthropods alone there are many examples of potential applications, including the nesting behavior of wasps as studied through the use of trap-nests which allow the nest-provision to be analyzed, the behavioral ecology of spider-hunting wasps, and the behavioral ecology of dung-beetles. These are all important aspects of applied entomology and ecology.

The study of flower-visiting is an important theme in ecology, with applied branches in pest control (including biological control of weeds, where natural enemies that attack flowers and prevent weed reproduction are particularly important) and crop production, where the pollination services of managed honeybees as well as wild pollinators is a topical subject. It has also been suggested that bees may be used in bee vectoring, defined as the use of managed pollinating bees to deliver beneficial microbial agents (fungi, bacteria and viruses) to flowering plants for the control of insect- or mite pests and suppression of plant diseases. These application areas could all potentially benefit from semantic enrichment and mediation of flower-visiting data in behavioral ecology studies, in which inferencing for the purpose of semantic interoperability could be used effectively.

In specialized groups, such as the *Rediviva* example described, where species have evolved strong mutualistic relationships, ontology design can allow specific behavioral ecology assertions to be made without detailed behavioral data. An example of such a design is defining the FV:*RedivivaFlowerProductSeekingEvent* class as a class of an ecological type, rather than a taxonomic class as its class name suggests, and confining the class to a specialized domain ontology. If a female *Rediviva* bee was found on an oil-plant we know that the bee was seeking floral oil even without observing the bee actually collecting oil. Similarly, in sub-Saharan Africa many species of bees in the genus *Lipotriches* specialize in collecting pollen from grasses (Pauly, 2014; Tchuenguem Fohouo *et al*., 2004). These examples therefore illustrate the kind of knowledge and approach that may be useful in future work in the area of modeling and semantic interoperability in behavioral ecology.

Further, a form of inferencing for knowledge discovery could be pursued through the identification of plants whose flowers have been visited by sequencing the Cytochrome Oxidase I gene (the ‘barcode of life’; Hebert *et al*., 2003) in the pollen grains collected from arthropods’ bodies. This would require the matching of reference gene sequences obtained from samples taken from plant specimens of known identity with the sequences obtained from pollen of unknown provenance. In some cases such inferencing would obviate the need for expert knowledge or behavioral ecology observations because an arthropod can only obtain floral pollen by visiting a plant’s flower. In other words, if the species of pollen has been identified by molecular means, even from a 10-year-old museum bee specimen, enrichment to the class of FV:*FlowerPollenTransferEvent* would be justified.

1. **Conclusion**

We have demonstrated that arthropod specimen data containing flower-visiting observations can be transformed into useful ecological information by representing behavioral ecology knowledge using ontologies, and that semantic enrichment and mediation can be automated. Assumptions were unavoidable in remedying the problem of missing data, but these were substantiated by accepted knowledge. Moreover, the knowledge model allowed the user to ignore enrichment that relied on assumptions.

Future work will involve building a new system layer to link the ecological interactions and their constituent events into a plant-visiting or flower-visiting ecological interaction network that will summarize the essential information in a way that is objective, enriched, standardized, consistent (i.e. semantically enriched and semantically mediated) and of a high quality (i.e. integrating expert- and implicit knowledge). Such a flower-visiting network could take the form of a directed acyclic graph representing a Bayesian network, which will, moreover, allow the user to model, and reason with, uncertainty in flower-visiting data.

The reported approach to developing the knowledge model and system implementation present opportunities for further addressing the challenge of analyzing data on complex ecological interactions using partial observations of biodiversity. The ontological reconciliation of the object-centric and event-centric views on biodiversity and ecology remains an unexplored area, where much expert morphological knowledge lies untapped.

Extending the system design for the objective of interoperating between distributed flower-visiting data-stores will make the system more useful to researchers. Future work could also focus on the strategy of using the complexity of animal behavior as an opportunity to enrich and refine the knowledge model through human input instead of seeing the incomplete model of behavior as a weakness. Because behavior is a unique and essential feature of biodiversity, more work in these areas could have a significant impact in semantic interoperability in BDEI.

**Acknowledgements**

We thank the JRS Biodiversity Foundation (http://www.jrsbdf.org) for financial support through a 2011 grant entitled ‘Improvement and Integration of Pollinator Biodiversity Information in Africa’. We are grateful to two anonymous reviewers, as well as Hamish Robertson and Bridgit Davis, who made suggestions to improve the manuscript.

**References**

Aderhold, A., Husmeier, D., Lennon, J. J., Beale, C. M., & Smith, V. A. (2012). Hierarchical Bayesian models in ecology: Reconstructing species interaction networks from non-homogeneous species abundance data. *Ecological Informatics*, *11*, 55–64. doi:10.1016/j.ecoinf.2012.05.002

Antezana, E., Kuiper, M., & Mironov, V. (2009). Biological knowledge management: the emerging role of the Semantic Web technologies. *Briefings in Bioinformatics*, *10*(4), 392–407. doi:10.1093/bib/bbp024

Arp, R., & Smith, B. (2008). Function, role, and disposition in Basic Formal Ontology. *Nature Precedings*, *1941.1*(713), 1–4. doi:10.1038/npre.2008.1941.1

Baskauf, S., & Webb, C. (2014). Darwin SW: Semantic Web terms for biodiversity data, based on Darwin Core. Retrieved March 01, 2014, from https://code.google.com/p/darwin-sw/

Bisby, F. A. (2000). The quiet revolution: biodiversity informatics and the internet. *Science*, *289*(5488), 2309–2312. doi:10.1126/science.289.5488.2309

Coetzer, W., Gon, O., Hamer, M., & Parker-Allie, F. (2012). A new era for specimen databases and biodiversity information management in South Africa. *Biodiversity Informatics*, *8*, 1–11. Retrieved from https://journals.ku.edu/index.php/jbi/article/viewFile/4263/4038

Collins, S. L., Bettencourt, L. M., Hagberg, A., Brown, R. F., Moore, D. I., Bonito, G., Delin, K.A.,

Jackson, S.P., Johnson, D.W., Burleigh, S.C., Woodrow, R.R. & McAuley, J. M. (2006). New opportunities in ecological sensing using wireless sensor networks. *Frontiers in Ecology and the Environment*, *4*(8), 402–407. doi:10.1890/1540-9295(2006)4[402:NOIESU]2.0.CO;2

Cooper, L., Walls, R. L., Elser, J., Gandolfo, M. a, Stevenson, D. W., Smith, B., Preece, J., Athreya, B., Mungall, C.J., Rensing, S., Hiss, M., Lang, D., Reski, R., Berardini, T.Z., Li, D., Huala, E., Schaeffer, M., Menda, N., Arnaud, E., Shrestha, R., Yamazaki, Y. & Jaiswal, P. (2013). The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant & Cell Physiology*, *54*(2), e1. doi:10.1093/pcp/pcs163

Daltio, J., & Medeiros, C. (2008). Aondê: An ontology Web service for interoperability across biodiversity applications. *Information Systems*, *33*(7-8), 724–753. doi:10.1016/j.is.2008.02.001

Deans, A. R., Yoder, M. J., & Balhoff, J. P. (2011). Time to change how we describe biodiversity. *Trends in Ecology & Evolution*, *27*(2), 78–84. doi:10.1016/j.tree.2011.11.007

Edwards, J. L., Lane, M. A., & Nielsen, E. S. (2000). Interoperability of biodiversity databases: biodiversity information on every desktop. *Science*, *289*(5488), 2312–2314. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11009409

Gallai, N., Salles, J., Settele, J., & Vaissiere, B. (2009). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics*, *68*(3), 810–821. doi:10.1016/j.ecolecon.2008.06.014

Gkoutos, G. V., Schofield, P. N., & Hoehndorf, R. (2012). The Neurobehavior Ontology: An ontology for annotation and integration of behavior and behavioral phenotypes. *International Review of Neurobiology*, *103*, 69–87.

Graybeal, J., Isenor, A. W., & Rueda, C. (2012). Semantic mediation of vocabularies for ocean observing systems. *Computers & Geosciences*, *40*, 120–131. doi:10.1016/j.cageo.2011.08.002

Hart, N. H., & Huang, L. (2012). Monitoring nests of solitary bees using image processing techniques. In *2012 19th International Conference on Mechatronics and Machine Vision in Practice, M2VIP 2012* (pp. 1–4).

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*, 313–321.

Horridge, M. (2011). A practical guide to building OWL ontologies using Protege 4 and CO-ODE Tools, Edition 1.3.

Jones, M. B., Schildhauer, M. P., Reichman, O. J., & Bowers, S. (2006). The new bioinformatics: integrating ecological data from the gene to the biosphere. *Annual Review of Ecology Evolution and Systematics*, *37*(1), 519–544. doi:10.1146/annurev.ecolsys.37.091305.110031

Kevan, P. G., & Baker, H. G. (1983). Insects as flower visitors and pollinators. *Annual Review of Entomology*, *28*, 407–453.

Krebs, J. R., & Davies, N. B. (1996). *Behavioral Ecology: An Evolutionary Approach* (Fourth Edi.). Sunderland, MA: Sinauer Associates.

Madin, J. S., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., & Villa, F. (2007). An ontology for describing and synthesizing ecological observation data. *Ecological Informatics*, *2*(3), 279–296. doi:10.1016/j.ecoinf.2007.05.004

Madin, J. S., Bowers, S., Schildhauer, M. P., & Jones, M. B. (2008). Advancing ecological research with ontologies. *Trends in Ecology & Evolution*, *23*(3), 159–68. doi:10.1016/j.tree.2007.11.007

Michener, W., Beach, J. H., Jones, M. B., Ludäscher, B., Pennington, D. D., Pereira, R. S., Rajasekar, A., & Schildhauer, M. (2007). A knowledge environment for the biodiversity and ecological sciences. *Journal of Intelligent Information Systems*, *29*(1), 111–126. doi:10.1007/s10844-006-0034-8

Michener, W., & Jones, M. B. (2012). Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology & Evolution*, *27*(2), 85–93. doi:10.1016/j.tree.2011.11.016

Michener, W., Porter, J., Servilla, M., & Vanderbilt, K. (2011). Long term ecological research and information management. *Ecological Informatics*, *6*(1), 13–24. doi:10.1016/j.ecoinf.2010.11.005

Midford, P. E. (2004). Ontologies for behavior. *Bioinformatics (Oxford, England)*, *20*(18), 3700–3701. doi:10.1093/bioinformatics/bth433

Moodley, D., Simonis, I., & Tapamo, J. (2012). An architecture for managing knowledge and system dynamism in the worldwide Sensor Web. International Journal of Semantic Web and Information Systems: Special issue on Semantics-enhanced Sensor Networks. *International Journal on Semantic Web and Information Systems*, *8*(1), 64–88.

Pauly, A. (2014). Les abeilles des graminées ou Lipotriches Gerstaecker, 1858, sensu stricto (Hymenoptera : Apoidea : Halictidae : Nomiinae) de l’Afrique subsaharienne. *Belgian Journal of Entomology*, *20*, 1–393.

Pauw, A. (2006). Floral syndromes accurately predict pollination by a specialized oil-collecting bee (*Rediviva peringueyi*, Melittidae) in a guild of South African orchids (Coryciinae). *American Journal of Botany*, *93*(6), 917–926. doi:10.3732/ajb.93.6.917

Peterson, A. T., Knapp, S., Guralnick, R., Soberón, J., & Holder, M. T. (2010). The big questions for biodiversity informatics. *Systematics and Biodiversity*, *8*(2), 159–168. doi:10.1080/14772001003739369

Rasmussen, C., & Olesen, J. M. (2000). Oil flowers and oil-collecting bees. *Det Norske Videnskaps-Akademi. I. Matematisk Naturvidenskapelige Klasse, Skrifter, Ny Serie*, *39*, 23–31.

Raven, P. H., Evert, R. F., & Eichhorn, S. E. (1986). *Biology of Plants*. New York: Worth Publishers, Inc.

Sala, A., & Bergamaschi, S. (2009). A mediator based approach to ontology generation and querying of molecular and phenotypic cereals data. *International Journal of Metadata, Semantics and Ontologies*, *4*(1/2), 85–92.

Scholes, R. J., Mace, G. M., Turner, W., Geller, G. N., Jürgens, N., Larigauderie, A., Muchoney, D., Walther, B.A., & Mooney, H. A. (2008). Toward a global biodiversity observing system. *Science*, *321*, 1044–1045. doi:10.1126/science.1162055

Tchuenguem Fohouo, F.-N., Pauly, A., Messi, J., Brückner, D., Tinkeu, L., & Basga, E. (2004). Une abeille afrotropicale spécialisée dans la récolte du pollen de graminées (Poaceae) : Lipotriches notabilis (Schletterer 1891) (Hymenoptera Apoidea Halictidae). *Annales de La Société Entomologique de France (Nouvelle Série)*, *40*(2), 131–143. Retrieved from http://zoologie.umh.ac.be/asef/contents.asp?action=detail&ARTID=385

Uschold, M., & Gruninger, M. (1996). Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, *11*(2), 1–63.

Villa, F., Athanasiadis, I., & Rizzoli, A. (2009). Modelling with knowledge: A review of emerging semantic approaches to environmental modelling. *Environmental Modelling & Software*, *24*(5), 577–587. doi:10.1016/j.envsoft.2008.09.009

Walls, R. L., Deck, J., Guralnick, R., Baskauf, S., Beaman, R., Blum, S., Bowers, S., Buttigieg, P.L., Davies, N., Endresen, D., Gandolfo, M.A., Hanner, R., Janning, A., Krishtalka, L., Matsunaga, A., Midford, P., Morrison, N., O´ Tuama, E., Schildhauer, M., Smith, B., Stucky, B.J., Thomer, A., Wieczorek, J., Whitacre, J., & Wooley, J. (2014). Semantics in support of biodiversity knowledge discovery: An introduction to the Biological Collections Ontology and related ontologies. *PLoS ONE*, *9*(3), e89606. doi:10.1371/journal.pone.0089606

Whitehead, V. B., & Steiner, K. E. (2000). Oil-collecting bees of the winter rainfall area of South Africa. *Annals of The South African Museum*, *108*, 143–277. Retrieved from http://biostor.org/reference/113424

Whitehead, V. B., Steiner, K. E., & Eardley, C. D. (2008). Oil collecting bees mostly of the summer rainfall area of southern Africa (Hymenoptera: Melittidae: *Rediviva*). *Journal of the Kansas Entomological Society*, *81*(2), 122–141.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., & Vieglais, D. (2012). Darwin Core: An evolving community-developed biodiversity data standard. *PLoS ONE*, *7*(1), e29715. doi:10.1371/journal.pone.0029715

Worboys, M. (2005). Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science*, *19*(1), 1–28. doi:10.1080/13658810412331280167