

# Big Data Opportunities in Biodiversity Informatics: A Functional Trait Perspective



Vaughn Shirey<sup>[1,2]</sup>, Vincent O'Leary<sup>[1,2]</sup>, Steven Dilliplane<sup>[2]</sup>  
 [1] Drexel University Department of Biodiversity, Earth, and Environmental Science (BEES)  
 [2] The Academy of Natural Sciences of Drexel University

THE ACADEMY  
 OF NATURAL SCIENCES  
 of DREXEL UNIVERSITY

## Introduction to Functional Biogeography Data Science

Advances in statistical “big data” approaches to ecology have been on the rise over the past decade. Functional biogeography provides an approach to modeling biological distributions based on quantifiable “traits” present on organisms (Violle, et al. 2014). Violle, et al. state that the “establishment of robust trait-environment relationships will help achieve one core goal of functional biogeography.” These trait environments are active in biological and environmental systems and interact over time with different consequences. The combination of these trait interactions (Figure 1) drives species distributions and community structure. Current approaches in machine learning use observations to describe and predict patterns, but lack the ability to explain predicted observations or propose practical models of biological mechanisms. This limits the progress on constructing a comprehensive *biological knowledge graph*.

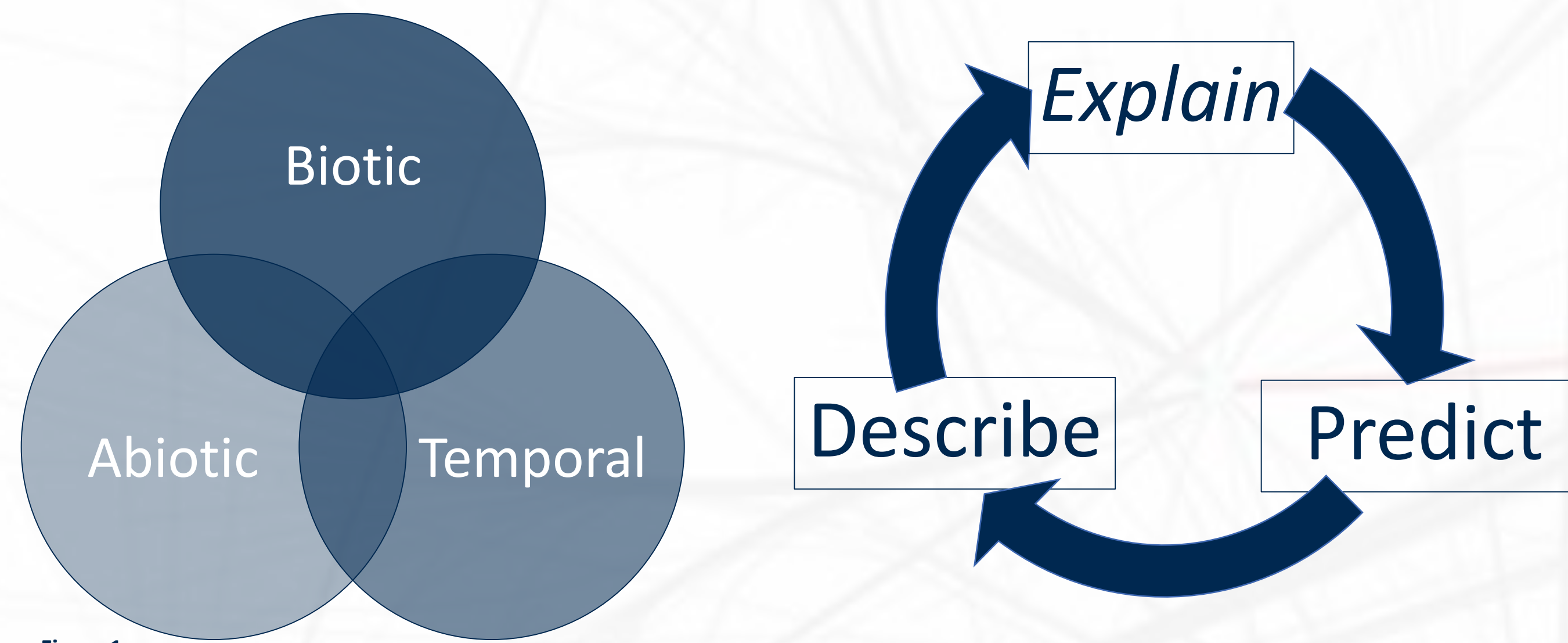


Figure 1

## Characterizing traits at variable scales

The increasing measurement and collection of a potentially intractable number of functional traits has created a massive volume of data concerning the phenotypic and behavioral attributes of a variety of species across spatial and temporal scales. In order to develop models that can explain and predict biogeographical trends, we must be able to define traits at different scales of study and interaction.

In Figure 2, a traditional biological hierarchy is presented to think about where traits can interact throughout a system. Functional traits represent a unique opportunity to synthesize knowledge and research goals across the multiple scales of biology from “-omics” and molecular scales through taxa, communities, and ecosystems (Figure 2). A similar hierarchy for non-living systems is needed in order to develop connections between scales and to the biological hierarchy.

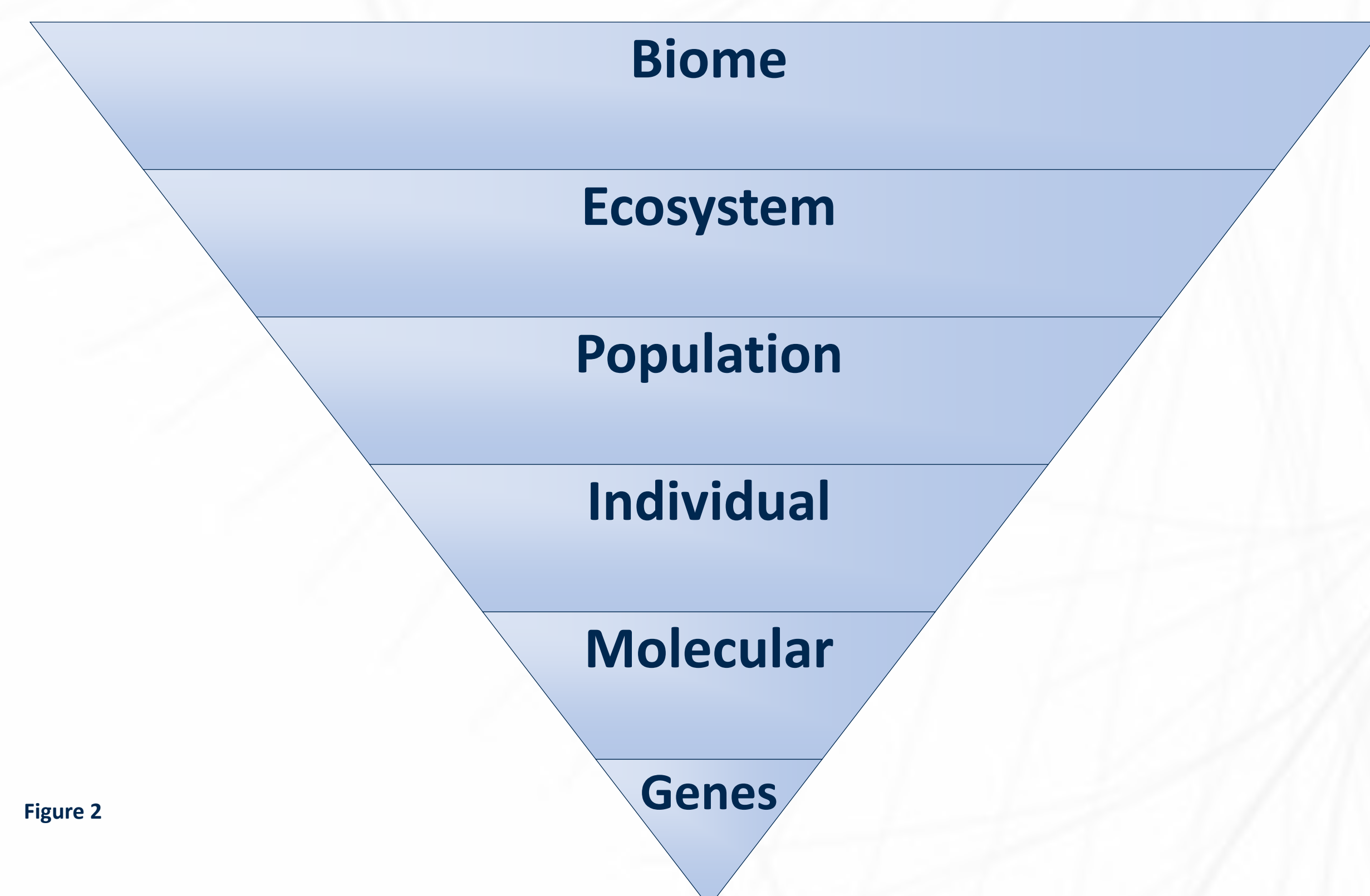


Figure 2

## Characterizing traits within systems

Functional traits are studied and expressed at variable scales (Figure 2), but can also be characterized in variable detail. The trait matrix below (Figure 3) describes the most frequently encountered types of biogeographic analysis. Our aim is to develop a robust framework for relating biotic and abiotic traits as described in the upper right quadrant.

biotic	<b>Functional</b>	Highlights a functional trait that allows for a set of organisms to live in a particular, named environment.  “Plants with a xylem diameter of .021 mm or larger are suited to live in deserts.”	Represented by relating the functional characteristics of life to functional characteristics of geography.  “Organisms that present traits x,y,z will be found to occupy environments that present traits x,y,z.”
	<b>Nominal</b>	Traditionally represented as high level classifications of geographic designation and taxonomic designation following some broad generalizations.  “Wolves live in forests.”	Traditionally represented as ecological niche modeling, predicting a species’ range based on tolerances for geospatially distributed functional parameters of geography.  “Species x is limited to range G where minimum temperature is T degrees and annual rainfall averages R inches.”
		<b>Nominal</b>	<b>Functional</b>
		abiotic	

Figure 3

## Designing an Ontology to Explore Trait Interactions

Representing the *biological knowledge graph* as a series of patterns, each instances of modular ontologies, suggests that a modular set of functional biogeography ontologies is needed as an interface ontology framework, bridging biotic traits and abiotic affordances, to model interactions among organisms and their environments.

A biological trait ontology includes measurable traits linked to both biological processes and a potential link to taxonomy. An environmental ontology includes habitat types and measurable variables specific to each type. For example, aquatic habitats would have many exclusive environmental variables compared to terrestrial habitats (Buttigieg, et al. 2016).

We present, for consideration and feedback, the case for utilizing functional trait data to develop a principled framework for integrating bio-ontology driven trait selection into the modeling and analysis of biological systems. In our test case, we illustrate the general selection of an aquatic habitat with a small sample of measurable variables that could interface with biological traits.

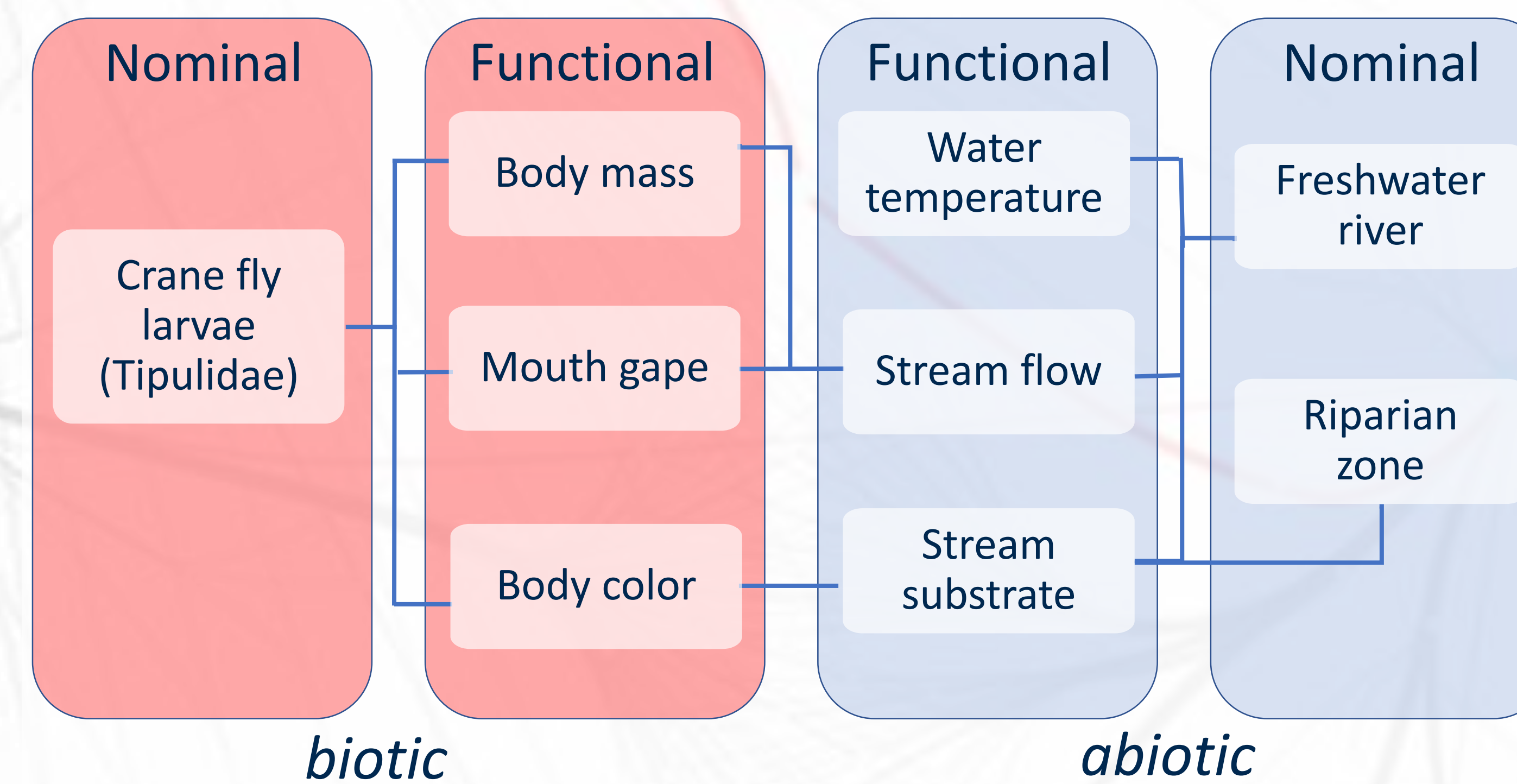


Figure 4

## Visualizing an n-Dimensional Hypervolume of Traits

Hutchinson’s *n-dimensional hypervolume* is a summation of all trait interactions we are trying to characterize across systems. The complexity of these networks suggests several key challenges hinder a direct “big data” approach to constructing a useful *biodiversity knowledge graph*. These challenges include 1) the need for an objective means of trait selection for modelling and 2) the ability to apply the resulting models across other spatial and/or temporal scales in more generalized contexts. By thinking about a framework or ontology to describe traits across systems, we can develop more thoughtful trait selection for models and apply our results to more contexts to increase understanding.

Trait (feature) selection and projection (mapping) provide principled techniques for managing the complexity inherent in attempt to understand high-dimensional data spaces. Generic network (graph) diagrams and parallel coordinates represent a spectrum of approaches to rationally organizing these spaces.

Hive diagrams have shown promise for visualizing complex networks in “-omics” biology/bioinformatics. Trait-selection could be reinforced through this model, providing a clear picture of potential traits selected to model biological systems based on their interactions. When these visuals are combined with the *ladder of abstraction* model by Victor, 2011 it provides robust exploratory tools. This is explored by a test case in Figure 8 below.

We hypothesize that the trait interactions are analogous to affordances and affordance signaling. We also propose that these interactions (within the context of a specific biological process) can be modelled over different spatial extents and timelines, including species lifespans or deep time.

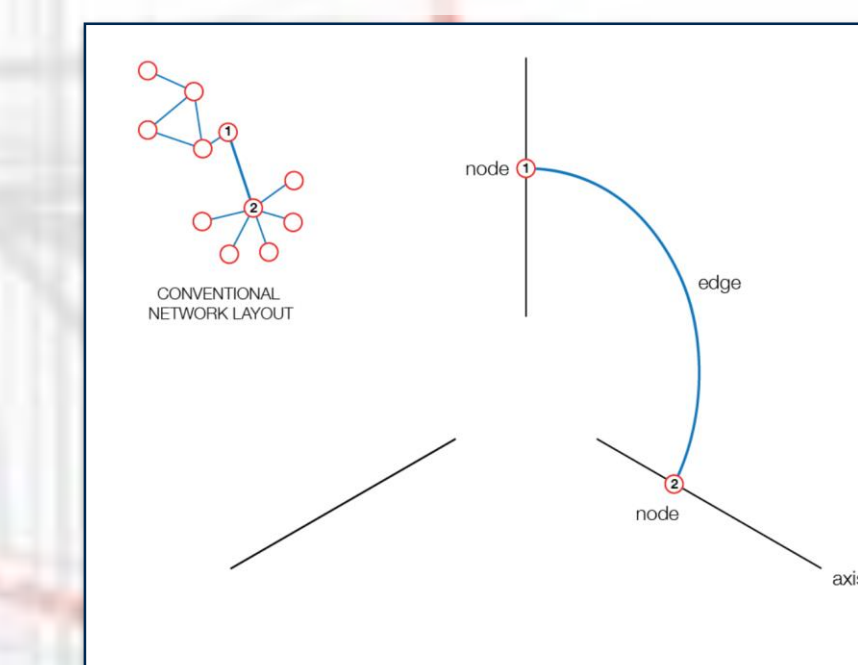


Figure 5

Figure 5: Translation of a network diagram to hive diagram with two nodes and a single edge (From Krzywinski 2010).

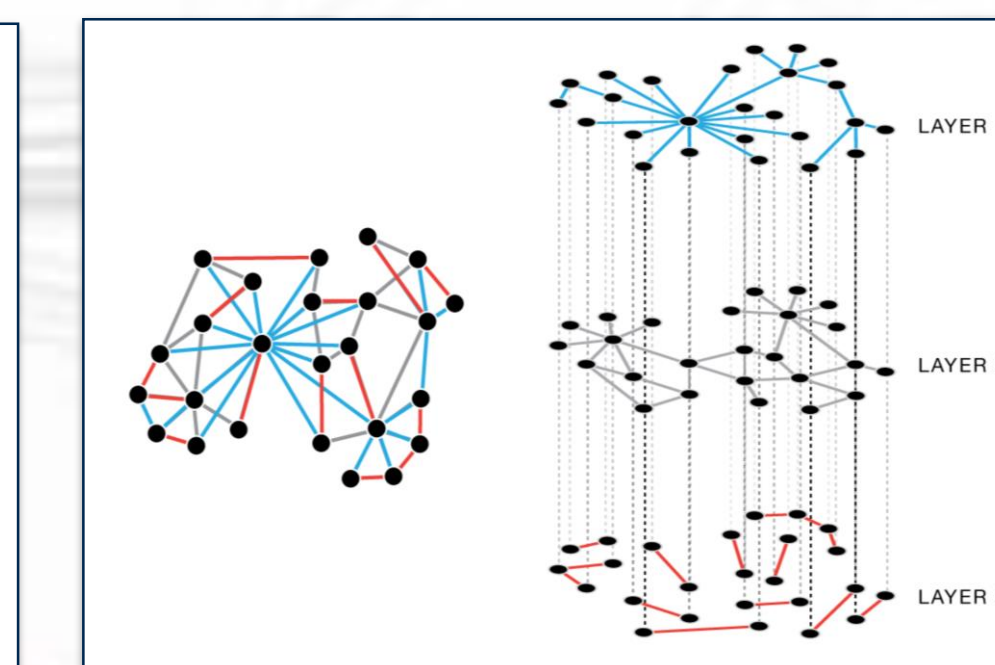


Figure 6

Figures 6/7: Multiple interacting systems under study can be directly compared using layered hive diagrams. (From Krzywinski 2010).

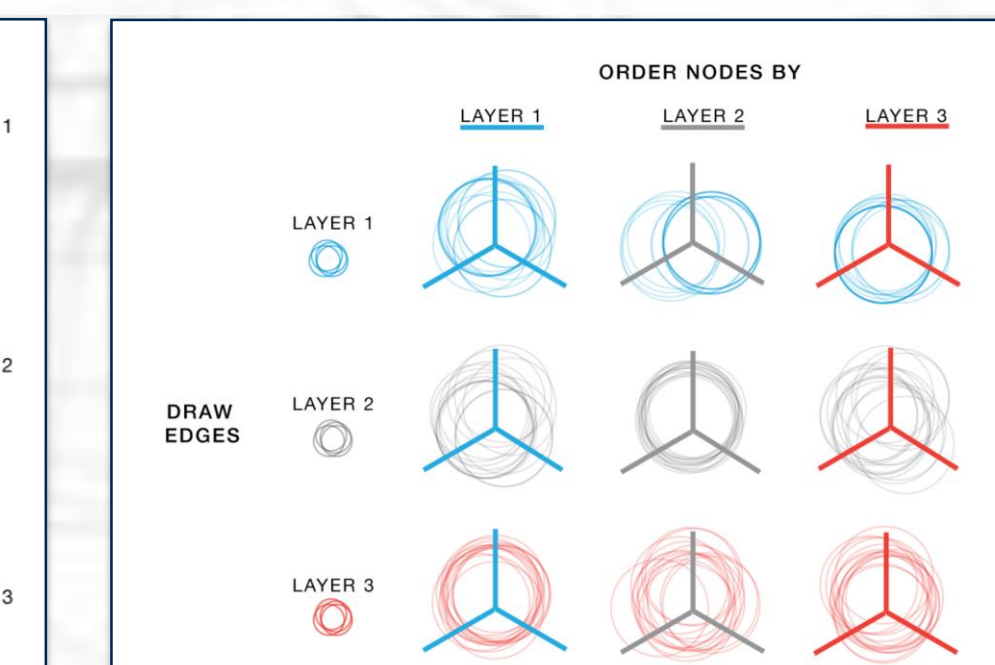


Figure 7

Figure 8: A test case exploring various traits under study interacting between abiotic, biotic, and temporal systems. The shape and structure would inform our understanding of the interactions.

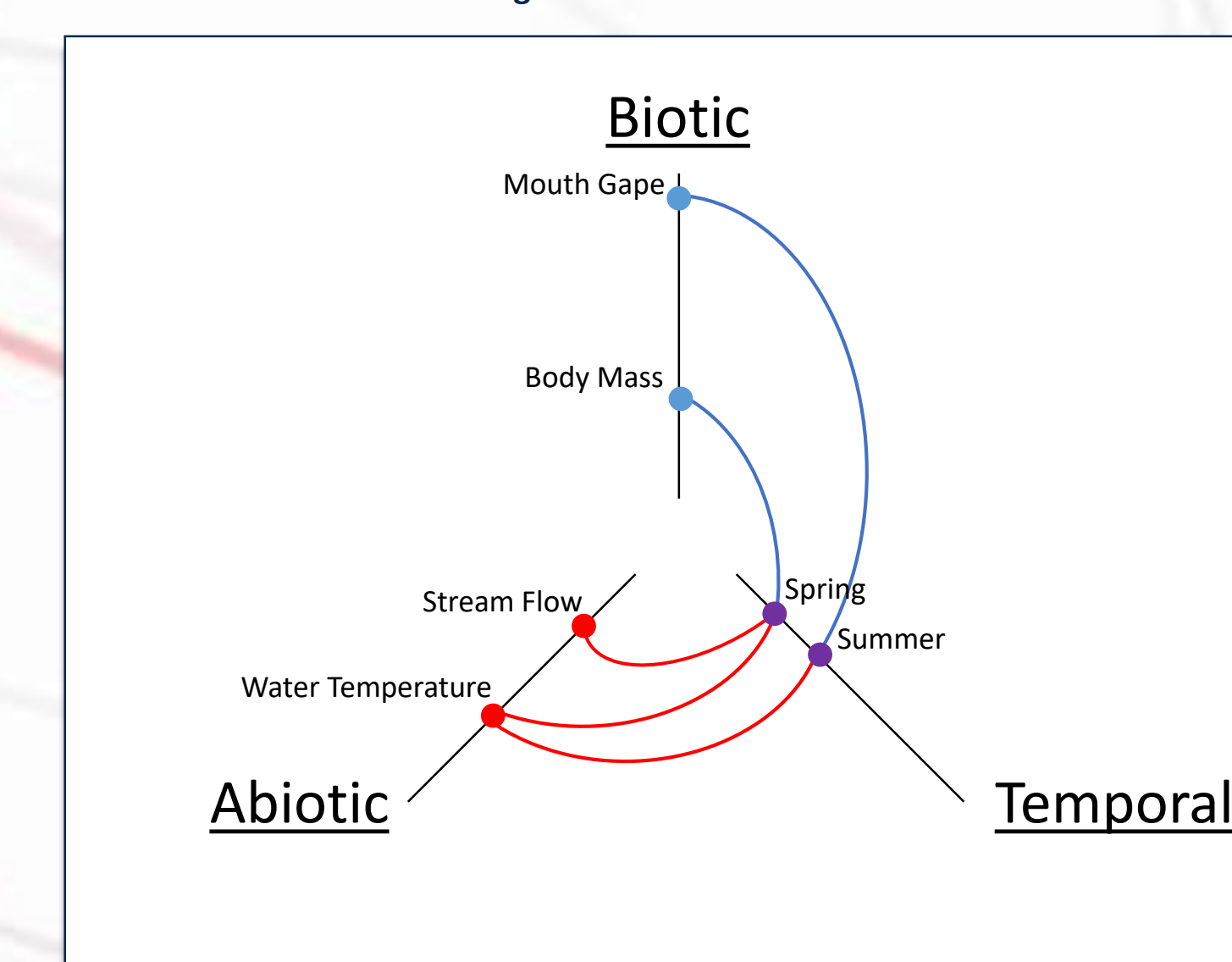


Figure 8

## Interested? Get Involved!

A project of this scale requires lots of collaboration to be successful. Please reach out to continue the discussion and be involved in future work as we continue to develop these ideas!

## References

All of the literature in this poster and more from our reading list can be found at [https://www.zotero.org/groups/big\\_data\\_opportunities\\_in\\_biodiversity\\_informatics](https://www.zotero.org/groups/big_data_opportunities_in_biodiversity_informatics) or by scanning the QR code with your phone.

