

Towards a Linked Biology – An integrated perspective of phenotypes and phylogenetic trees

Eduardo Miranda¹, Anaïs Grand², Régine Vignes Lebbe³ and André Santanchè¹

¹ Institute of Computing – State University of Campinas, Brazil

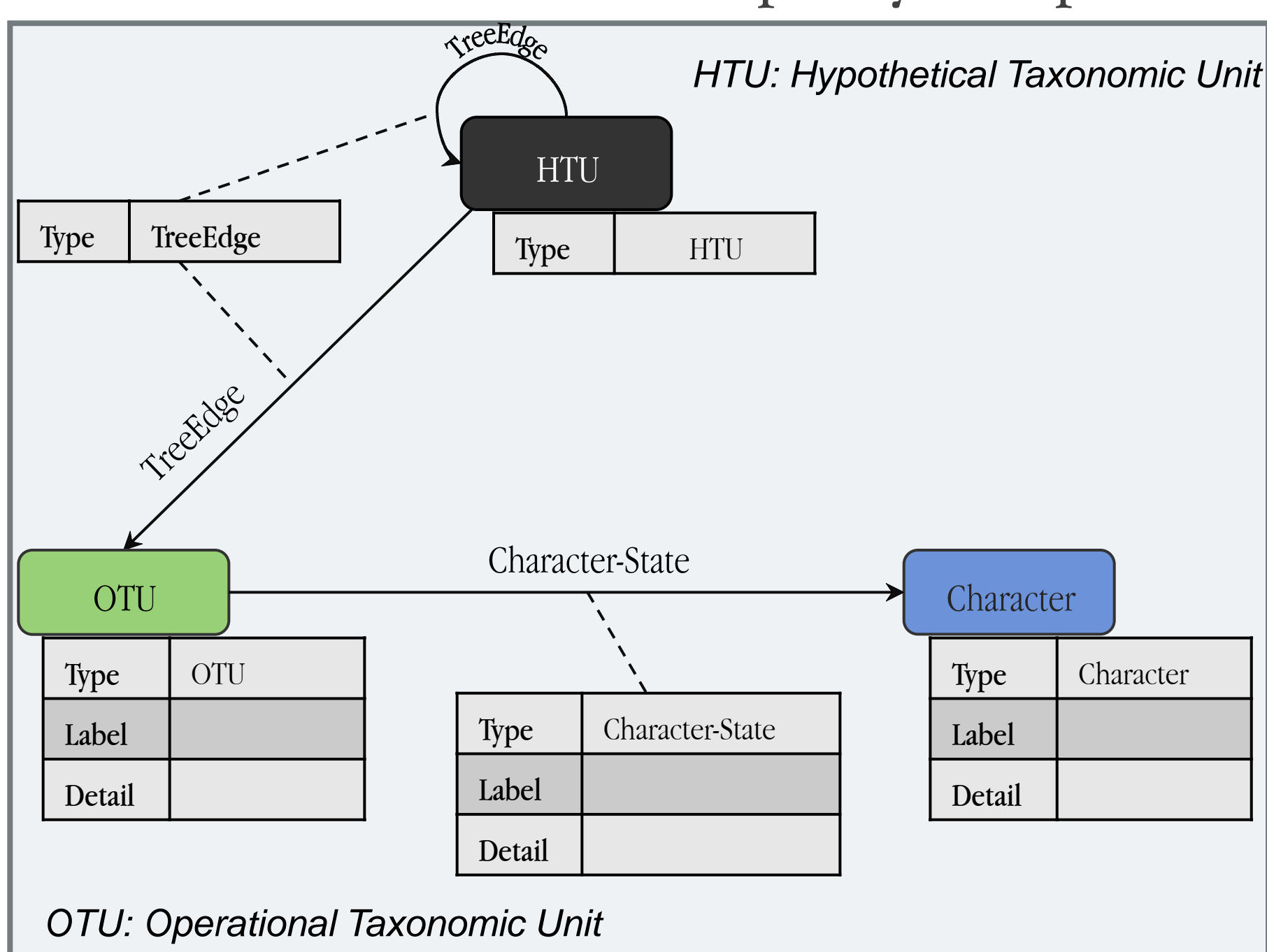
² Museum national d'histoire naturelle, CR2P - UMR 7207 CNRS/MNH/Univ Paris 06, France

³ UPMC Univ Paris 06, CR2P - UMR 7207 CNRS/MNH/Univ Paris 06, France

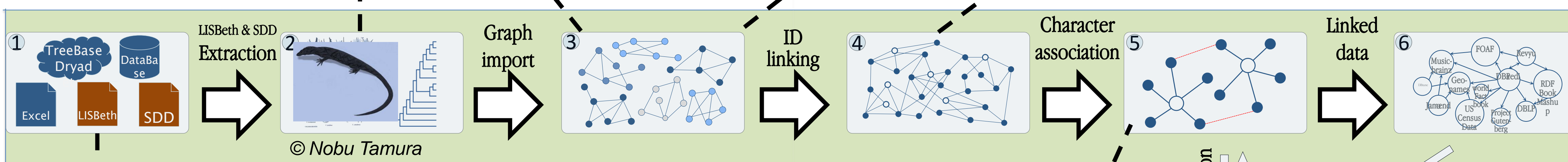
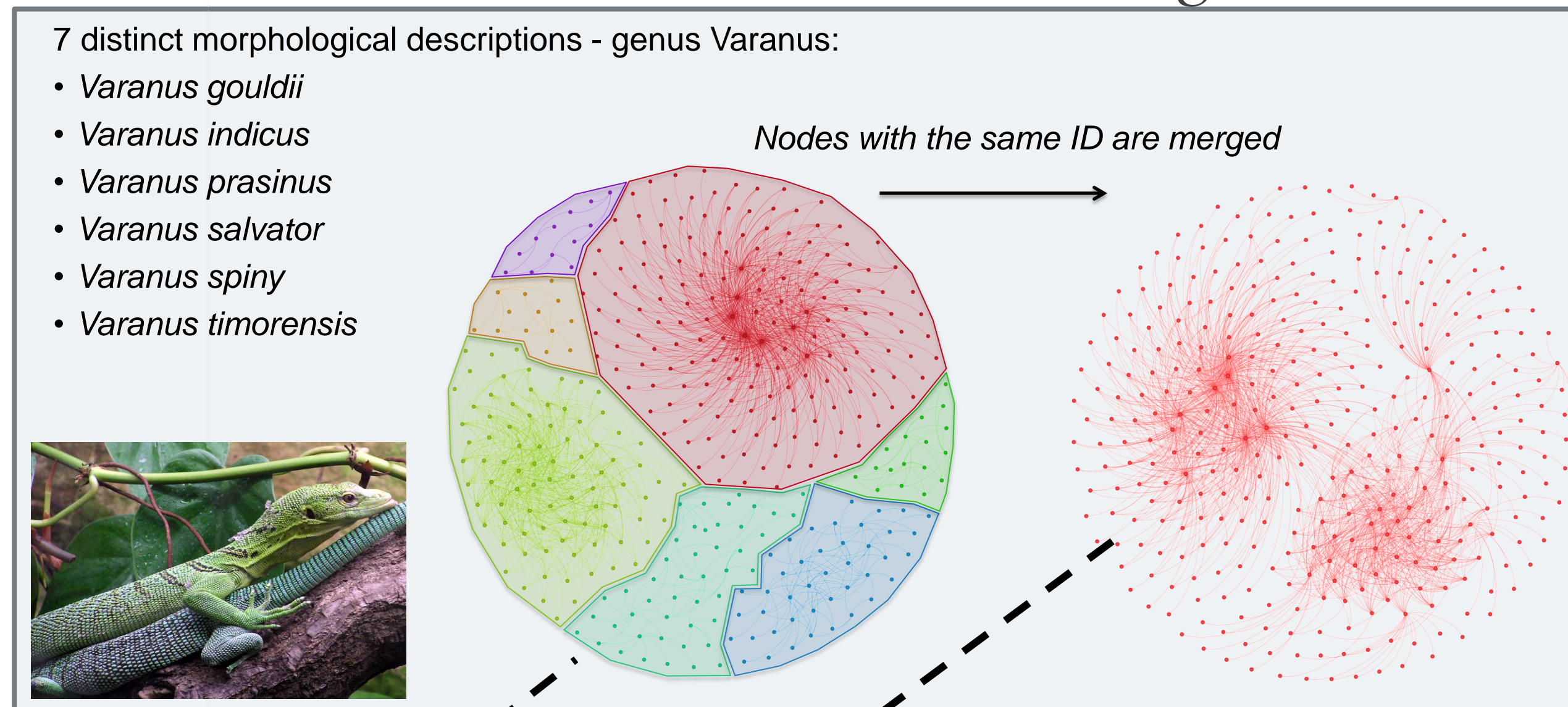
Background

A large number of studies in biology, including those involving phylogenetic tree reconstruction, result in the production of a huge amount of data – e.g., phenotype descriptions, morphological data matrices, etc. Biologists increasingly face a challenge and opportunity of effectively discovering useful knowledge by crossing and comparing several pieces of information, not always linked and integrated.

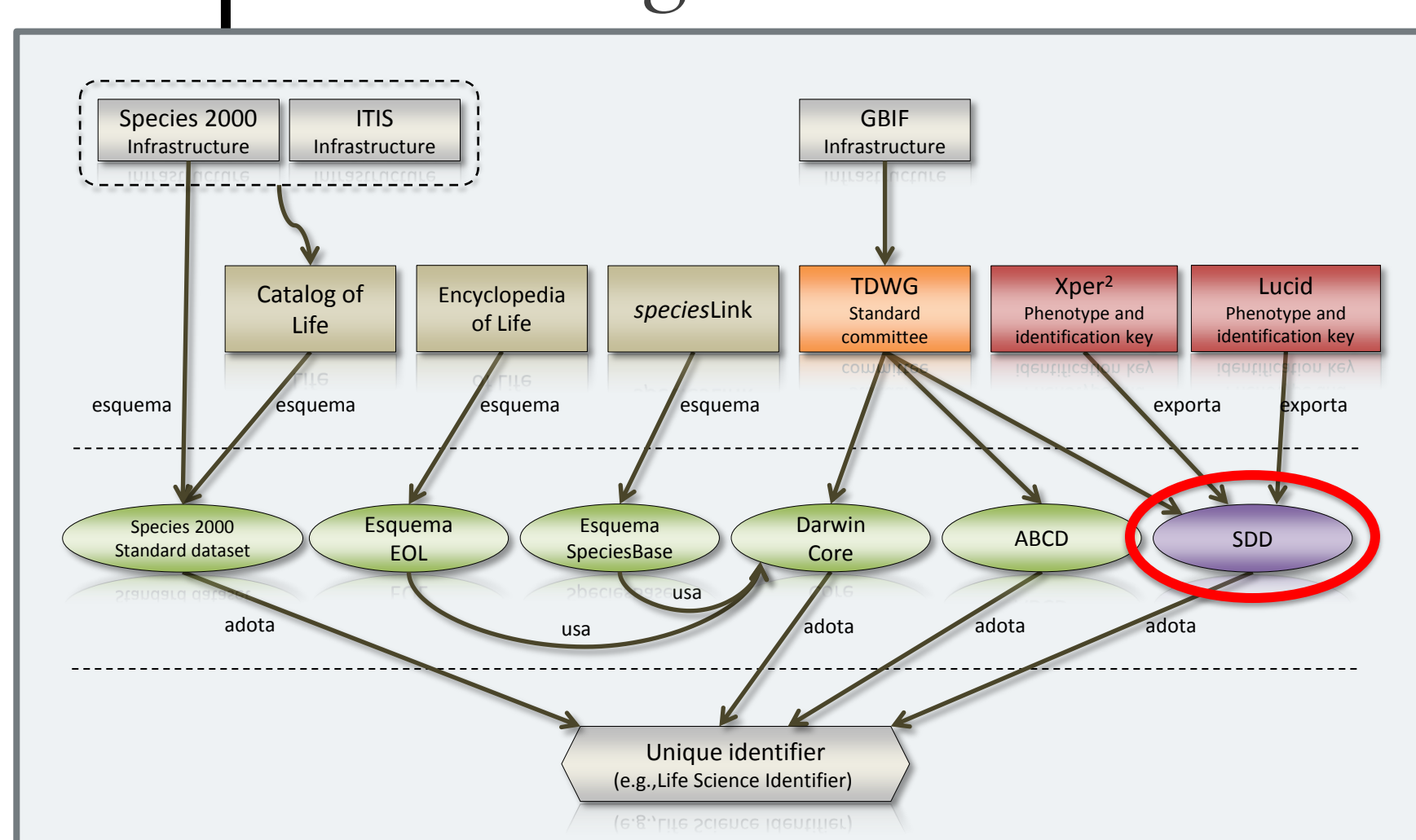
Unified Model as Property Graph



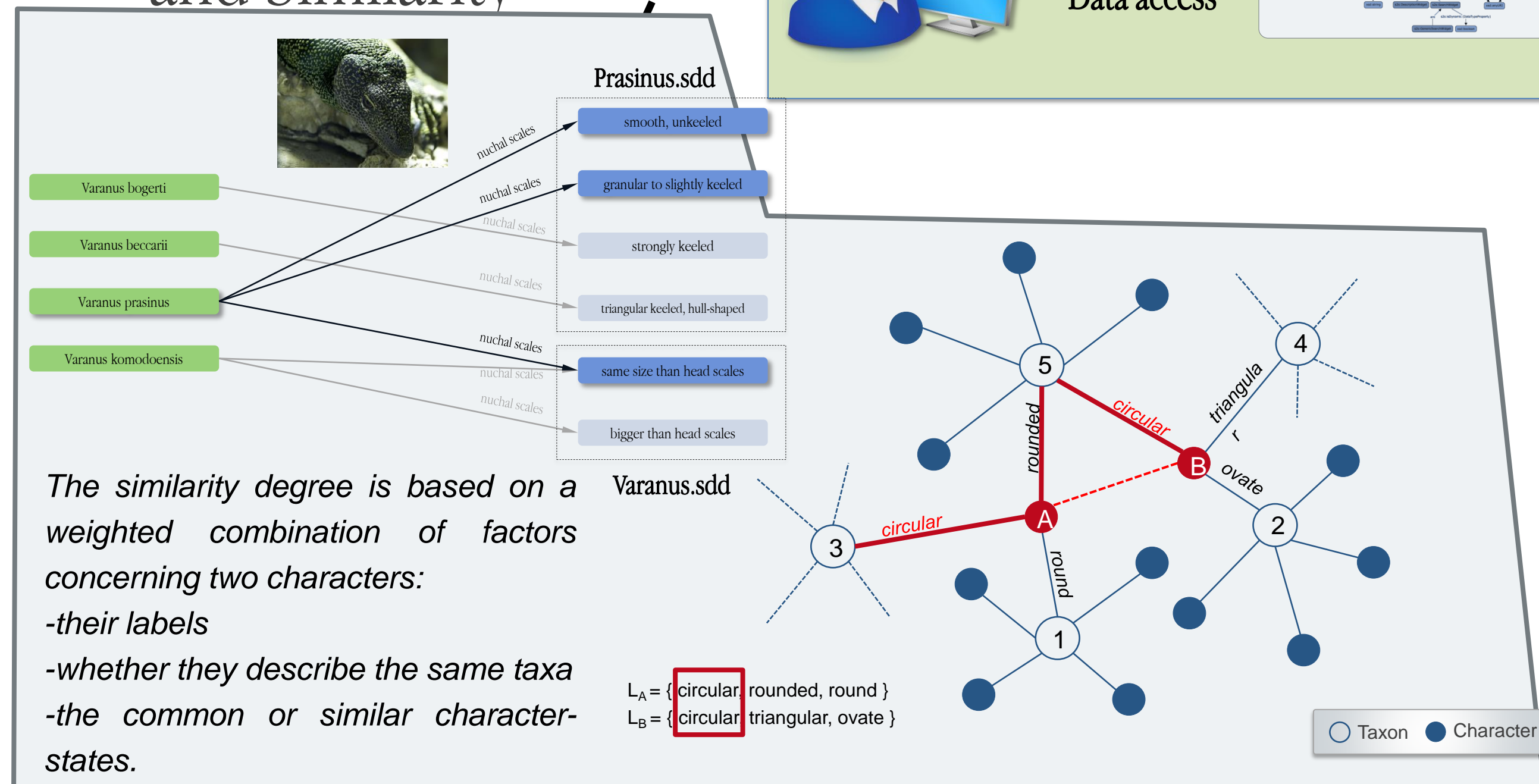
Varanus Identifier Linking



Heterogeneous data sources



Character association and Similarity



Conclusion

Our unified model enabled us to discover and make explicit the potential semantics raised by linking previously unconnected information. We developed a similarity index to visualize how close related are two given characters and some preliminary results are presented. This index has the potential to indicate the recurring use of the same character in different studies and might support biologists to understand and detect correlation between characters.

Acknowledgments. Work partially financed by (CNPq 138197/2011-3), the Microsoft Research FAPESP Virtual Institute (NavScales project), CNPq (Mu- ZOO Project and PRONEX-FAPESP), INCT in Web Science(CNPq 557.128/2009- 9) and CAPES-COFECUB, as well as individual grants from CNPq.