

# FLOPO - An Ontology for the Integration of Trait Data from Digitized Floras and Plant Image Collections

Claus Weiland<sup>1</sup>, Robert Hoehndorf<sup>2</sup>, George Gosline<sup>3</sup>, Quentin Groom<sup>4</sup>,  
Thomas Hamann<sup>5</sup>, and Marco Schmidt<sup>1</sup>

<sup>1</sup> Senckenberg Biodiversity and Climate Research Centre Frankfurt

<sup>2</sup> King Abdullah University of Science and Technology, Thuwal

<sup>3</sup> Royal Botanical Gardens Kew

<sup>4</sup> Botanic Garden Meise

<sup>5</sup> Naturalis Biodiversity Center Leiden

Knowledge about organismic traits is useful for many different studies in environmental sciences. For example, traits can be used to study the functional diversity of plant communities and competition between species. Furthermore traits of underutilized plants are an important resource for agricultural usage and phytomedical applications. There are large biodiversity data collections which are a potentially rich source of such plant trait data:

- Floras describe plant traits with a focus on morphology or habitus relevant for species identification in addition to other characteristics (e.g. health applications).
- Complementary topic databases, for example focusing on geographical distributions like Senckenberg’s West African Plants [1], provide large image collections which allow additional trait and habitat data to be extracted.

However, two key limitations in systematically analyzing trait data are the lack of a standardized vocabulary for the described traits and difficulties in extracting structured information from different knowledge sources. We have developed the Flora Phenotype Ontology (FLOPO [2]), an ontology for describing traits of plant species found in Floras.

Our methodological approach was to use NLP to extract entity-quality relationships from digitized taxon descriptions in Floras [3]. On this basis, we used a formal approach based on phenotype description patterns and automated reasoning to generate the FLOPO, which comprises more than 24000 phenotype and trait classes. We are currently applying deep learning approaches (convolutional neural networks) to extract additional plant trait data from image data like herbarium scans and photographs.

## References

1. <http://www.westafricanplants.senckenberg.de>
2. <http://aber-owl.net/ontology/FLOPO>
3. Hoehndorf, R. et al.: Journal of Biomedical Semantics 7:65 (2016)