**Data knowledge discovery (WP1)**

*Evolutionary biology-inspired semantic algorithms*: Most studies of data discovery focus on advanced analytics functions to reveal insights, ignoring the discovery of data-source heterogeneity almost completely. Currently, only a few databases are semantically annotated from many data-sources (e.g., gene ontology database, COVID-19). Ontology development is time-consuming and requires expert knowledge. It is also ideally paired with data-driven research that iteratively checks the soundness of the ontology as it simultaneously seeks discovery. Thus, software tools for mapping and linking the terms between different ontologies accounting for many data-sources are still not in place [6, 4].

*Going beyond* ROBHOOT will go beyond state-of-the-art to implement evolutionary-biology inspired semantic algorithms. We will explore insertions and deletions, different types of recombination and crossover, and other



Figure 1: **Discovery in evolutionary diversification-inspired federated networks**. ROBHOOT target knowledge discovery when heterogeneous groups of species, humans and technologies share ecosystem resources for a sustainable knowledge-inspired society: It introduces three science-enabled technologies: Evolutionary biology-inspired semantic algorithms for ROBHOOT v1.0 (data knowledge discovery, red), eco-evolutionary diversification-inspired AI models for ROBHOOT v2.0 (causal knowledge discovery, green), and evolutionary neural diversification-inspired federated networks for ROBHOOT v3.0 (discovery in federated networks, blue). ROBHOOT uses the sustainability of the Seas case study in federated networks to offer a compact open-source technology with full reproducibility, automation, visualization and reporting for an open citizen science.

evolutionary-based functions to find datatype properties from ontologies and raw-data from non-semantic databases. ROBHOOT will explore algorithms to gain understanding of the replicability of data heterogeneity contrasting different evolutionary algorithms.

**Causal knowledge dicovery (WP2)**

*Eco-evolutionary diversification-inspired AI algorithms*: Causal, explainable or interpretable discovery from observable data has been extensively studied (refs ++). Many of these studies have used symbolic reconstruction of equations by symbolic regressions or evolutionary methods (refs +++). However, a common gap throughout much of the literature is that of reconstruction from partially known models where the parameters represent eco-evolutionary and diversification processes from where explainability of the data can be done efficiently. The classical view on biology-inspired information processing technologies is to consider plasticity without structural changes, or without diversification among many interacting components (refs +++). Recent experimental evolution studies shows that rapid trait changes with new information processing capabilities is far more complex (refs +++). For example, eco-evolutionary dynamics strongly affect feedbacks between ecological and evolutionary processes, which in turn influences trait changes to open new properties with new information capabilities (refs +++). Furthermore, recent studies suggest that the interplay between trait dimensionality and adaptation is key to understand the emergence of new traits and information processing abilities to elaborate new discovery computation strategies in ecosystems (refs+++).

*Going beyond* ROBHOOT will, for the first time, employ eco-evolutionary diversification-inspired solutions to implement AI process-based methods to represent spatiotemporal causal inference in systems containing large heterogeneity and dimensionality (Figure 2). Eco-evolutionary diversification-inspired models will be extended to deep process-based learning networks including trait and interactions as evolutionary changes to understand patterns in these systems. The search for causal knowledge discovery will be applied to the data knowledge discovery generated in WP1 for the sustainability of the Seas. This database started in 1965 and currently co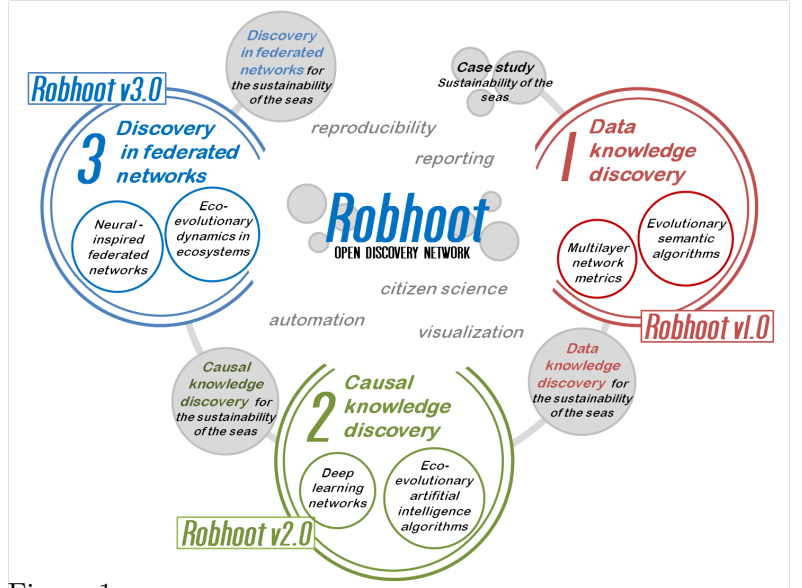ntains 9 mill. entries, 1612 species (i.e., 50 variables and