

# Extracting information from scientific articles to compile data on ecological impacts of biological invasions

Franck Courchamp<sup>1,\*</sup>

<sup>1</sup> Ecology, Systematics & Evolution, CNRS – University of Paris Saclay, France

## Abstract

Despite the importance of the ecological impacts of biological invasions, our understanding remains limited due to their diversity, which prevents any generalization. There are indeed at least 16 different types of ecological impacts, from local extinction to habitat degradation, and from trophic web modification to ecosystem service loss, and these are difficult to compare and impossible to compile. It is therefore crucial to develop better methods for assessing these impacts globally and understanding their underlying mechanisms. The aim of our project is to fill this critical gap by carrying out an integrated assessment of these impacts, which we will achieve by standardizing and quantifying all these impacts through a unified metrics. This will be obtained through the application of this metric to all known impact that have been published in the literature, to populate a database of quantified ecological impacts and their various descriptors. In order to achieve this, we need to obtain a dataset of all published studies on ecological impacts in the scientific literature, and to extract information related to these impacts from the pdf of these studies. As there is an estimated 25,000 such studies, we can only hope to achieve this through AI and AL.

## Keywords

Biodiversity, Biological Invasions, text information extraction

## 1. Introduction

Biological invasions are a major ecological and environmental issue of the 21<sup>st</sup> century, with significant impacts on native ecosystems, biodiversity, and human societies. Biological invasions can be defined as is the introduction of alien species to a new area where they establish and spread and have an impact in the invaded area. The importance of biological invasions is highlighted by their ubiquity and diversity. Invasive alien species can be found in almost every ecosystem on Earth, from terrestrial to freshwater and marine environments. The diversity of invasive alien species is also remarkable, ranging from microorganisms to vertebrates, and from plants to animals. In addition, the impacts of invasive alien species can be highly variable, depending on the ecological context, species interactions, and other factors. In addition to economic and sanitary impacts on human societies, biological invasions are indeed responsible for ecological impacts that include biodiversity loss, species extinctions, habitat degradation and alteration of ecosystem processes.

Despite the importance and ubiquity of biological invasions, our understanding of their ecological impacts remains limited, mostly due to all these levels of diversity that preclude generalizations. There is a clear need to develop better methods to assess the ecological impacts of invasive species, as well as to understand the mechanisms that drive these impacts. This information is critical for effective management and control of invasive species, as well as for the conservation of native biodiversity and ecosystem services. It is also key for communication purposes, as the

---

*3rd Workshop on Augmented Intelligence for Technology-Assisted Review Systems (ALTARS 2024): Evaluation Metrics and Protocols for eDiscovery and Systematic Review Systems. March 28, 2024, Glasgow, UK.*

✉ franck.courchamp@cnrs.fr

ORCID 0000-0001-7605-4548



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

diversity and variety of impact types hinders a good understanding, awareness and support of biological invasions as a major, global driver of environmental degradation.

Unifying the various types of impacts into a single metric will allow not only to build a strong framework that will lead to a better understanding and awareness of biological invasions, it would also allow impact quantifications and therefore permit compilations and comparisons across regions, taxonomic groups, impacted sectors and time from the published literature on this field. As the metric has now been obtained, we need to apply it to the studies reporting the ecological impacts, in order to transform each of them into our metric.

There are allegedly about 25,000 studies reporting impacts of given biological invasions of different species into various ecosystems and regions in the world, and therefore the authors of this projects are turning towards automated systems to fulfill this daunting task of extracting all the described impacts from these papers into a dedicated database, InvaPact. The team working on this project has achieved a similar process through the standardization and collection of economic costs of biological invasions, with the conception and population of the InvaCost database that regroups over 13,500 costs, each with 65 descriptors. This [InvaCost database](#) could be used as a training system for a system aiming at doing a similar exercise on ecological impacts in InvaPact.