From Shells in House Cabinets to Structured Data for Research:
The mobilization of frozen biodiversity data in Italy

Arianna Giannini, Marco Oliverio

Department of Biology and Biotechnology *Charles Darwin*, La Sapienza University, Rome, Italy

Introduction

In the era of big data, occurrences are increasingly disconnected from actual specimens, leading to a potential loss in information.

In Italy – as in other countries – many specimens of invertebrates are stored in private house cabinets. We call this type of data **frozen**.

Why private collections data?

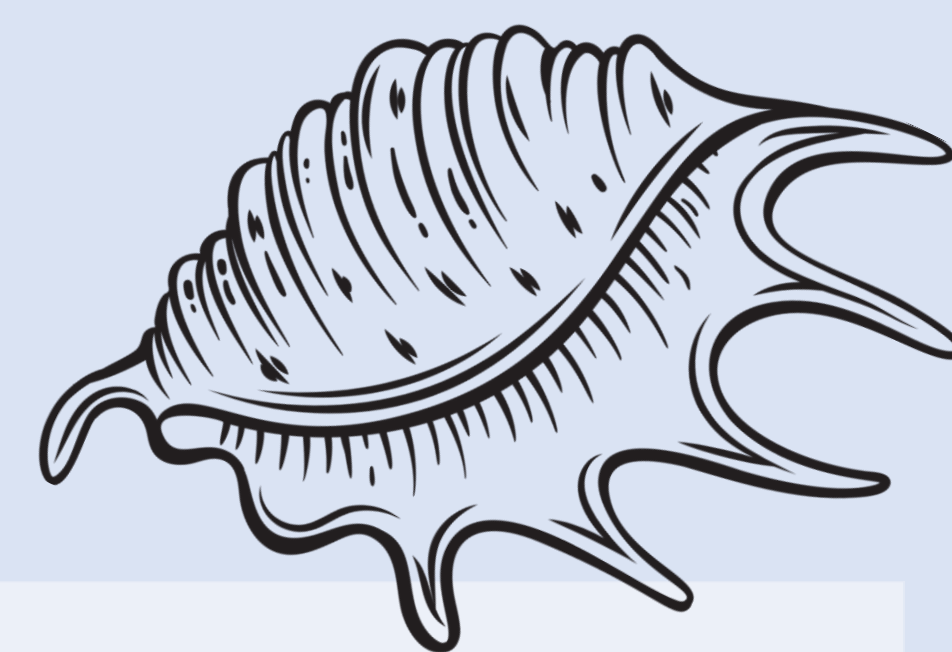| Source	Public collections	Private collections	Structured CS* projects	Other observation data
Data type	Specimen based	Specimen based	Observation based	Observation based
Quality	↑	↑	↓	↓
Accessibility	↔	↓	↔	↓
Ancillary data	↑	↑	↔	↔
Usability	↑	↑	↓	↓

*Citizen-Science

- High taxonomic and geographic quality;
- Ancillary data can be acquired at any moment;
- They are usable for both taxonomic, phylogenetic and ecological studies.

The real struggle is the *accessibility* of the data.

▶ **Mobilizing** Italian frozen data could **enhance** biodiversity research and management.

Methodology

Marine molluscs were chosen as pilot taxon based on taxonomy and conservation criteria.

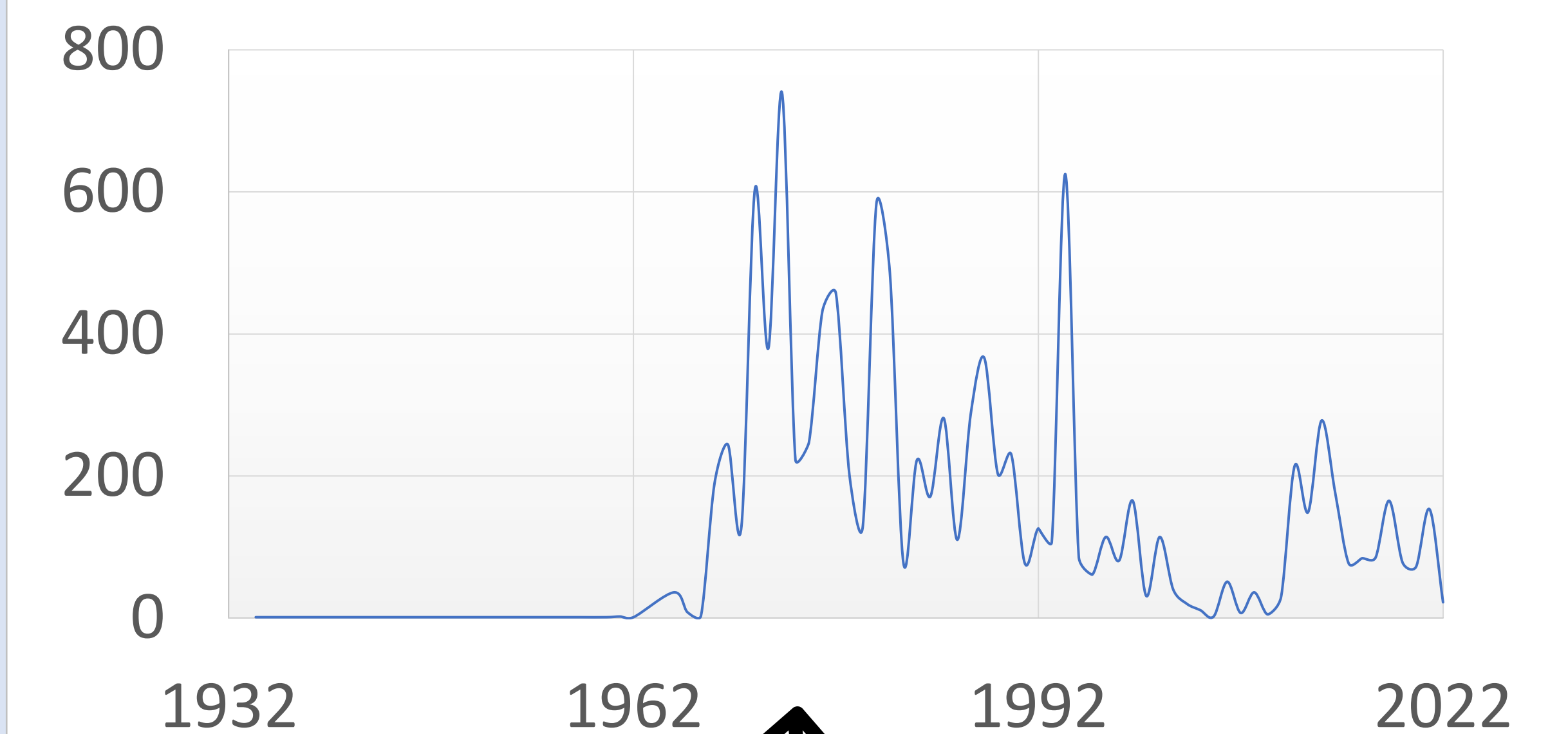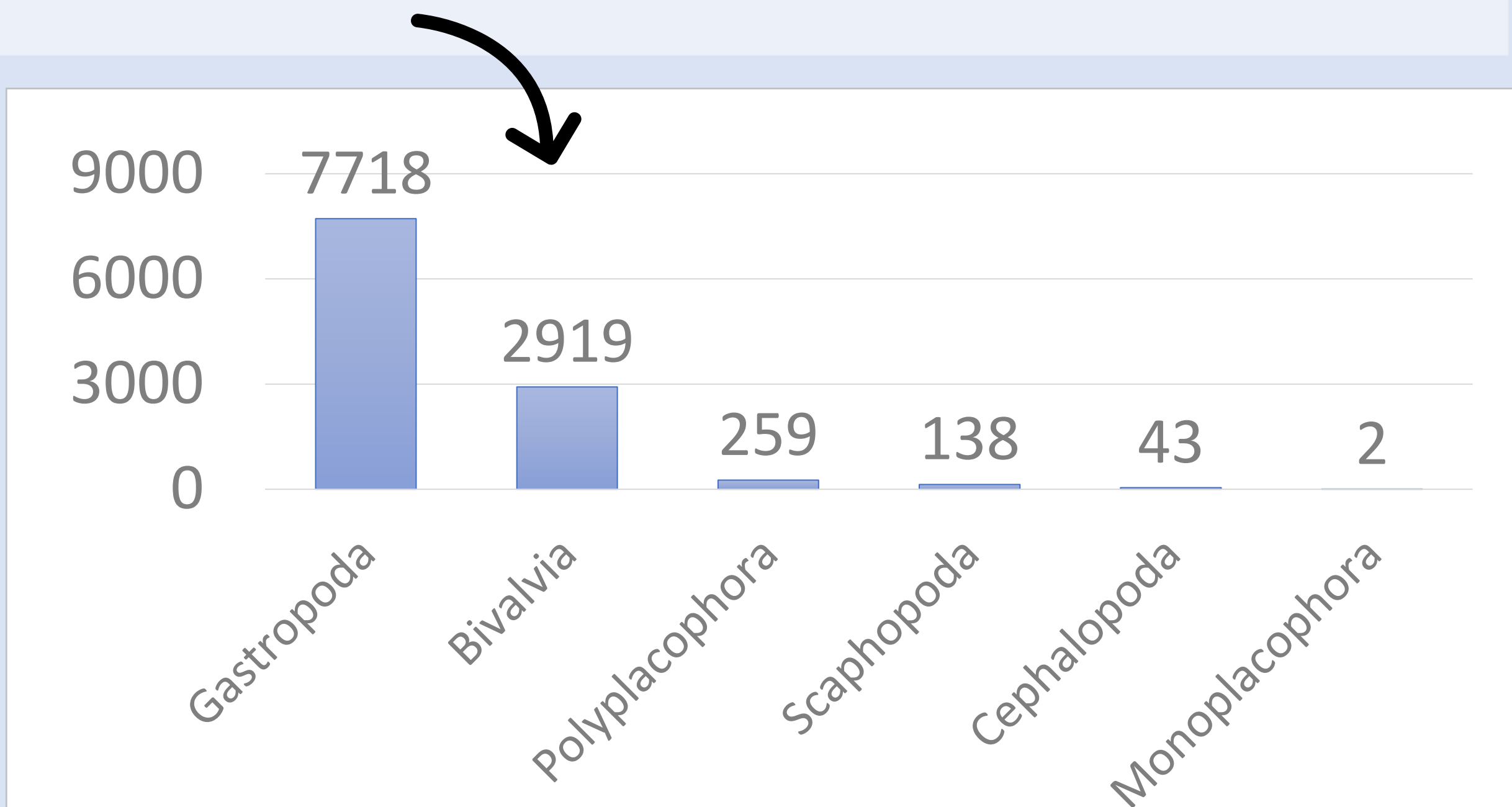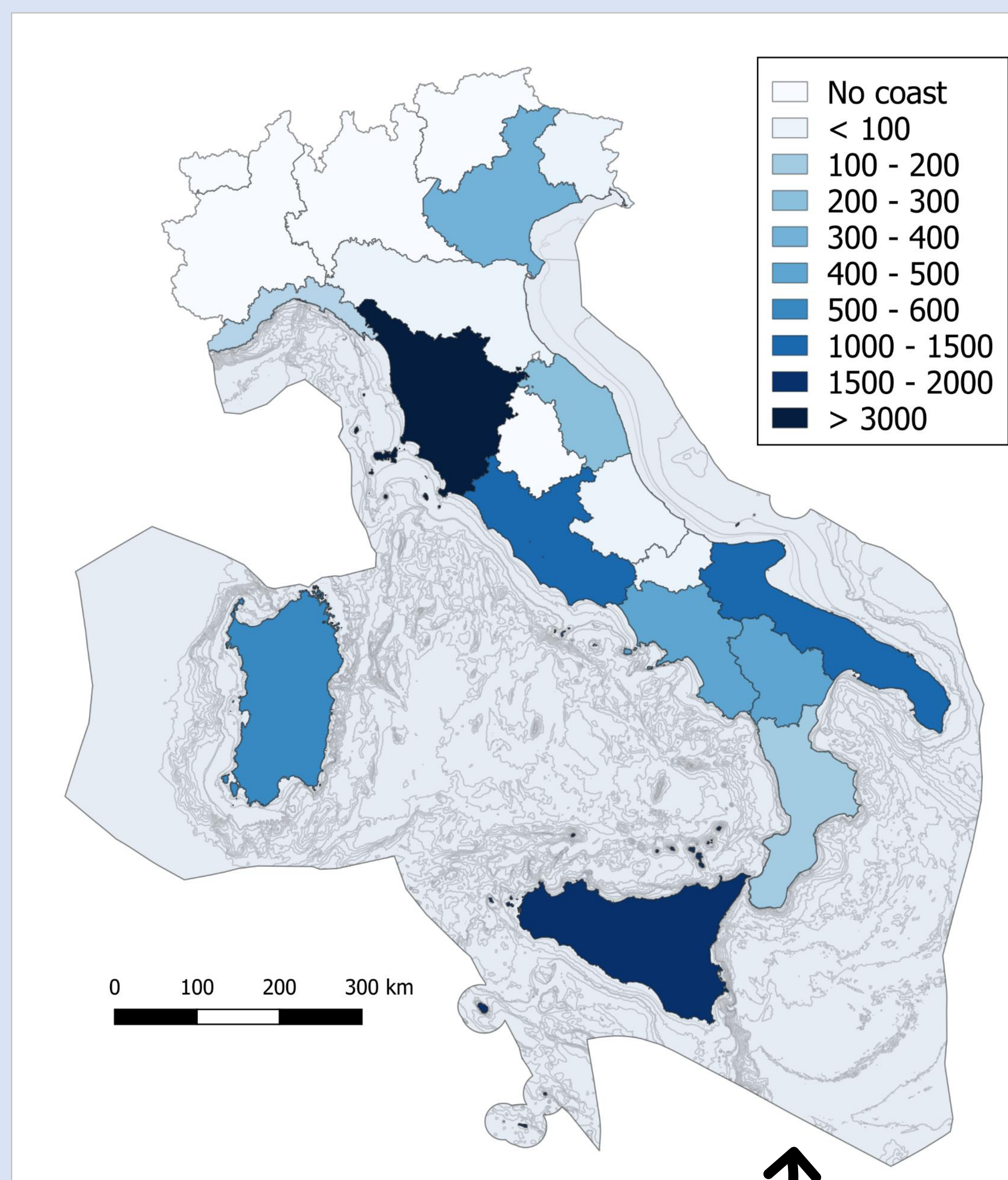A collector agrees to cooperate. → We send them a simple 21-fields Excel spreadsheet to fill out and a guide.

When data are ready we convert descriptive localities into coordinates and we standardize taxonomy and terms.

Collectors aren't always scientists! Requesting data simply and clearly without using specialized terms is important.

Preliminary results

About species

More than 11000 specimens were databased, coming from 3 different private collectors. Gastropods are the most sampled taxonomic group. In addition, the database contains records of 16 alien species for Italian waters.

Bar chart values: Gastropoda 7718, Bivalvia 2919, Polyplacophora 259, Scaphopoda 138, Cephalopoda 43, Monoplacophora 2

Map legend:
No coast
< 100
100 - 200
200 - 300
300 - 400
400 - 500
500 - 600
1000 - 1500
1500 - 2000
> 3000

0 100 200 300 km

Time series chart: 1932 – 1962 – 1992 – 2022 (values 0 to 800)

A strong geographic bias

Much of the dataset consists of specimens collected along the Tuscan and Sicilian coasts. Together, they make up more than 40% of the dataset. The Adriatic Sea turns out to be the least sampled (compared with the other Italian Seas).

A matter of time

The collecting time span of the dataset is from 1934 to 2022, but, although the first sampling event was conducted in 1934, no specimens were collected after this date until 1956. It was not until the 1960s that the number of samplings intensified.

Next steps

- Reach more collectors and expand the search to new taxa.
- Explore the drivers that create the sampling bias.

CHECK THE ABSTRACT!