



Dalmatian pyrethrum spatial records quality from Croatian Open Databases

Filip Varga^{1,2}, Martina Grdiša^{1,2}, Toni Nikolić³, Emanuel Guberović⁴, Ivana Bosnić⁴, Frederika Welle Donker⁵, Dragica Šalamon¹

¹University of Zagreb, Faculty of Agriculture, Svetošimunska 25, Zagreb, Croatia (Contact: fvarga@agr.hr)

²Centre of Excellence for Biodiversity and Molecular Plant Breeding, Svetošimunska 25, Zagreb, Croatia

³University of Zagreb, Faculty of Science, Department of Biology, Rooseveltov trg 6, Zagreb, Croatia

⁴University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, Zagreb, Croatia

⁵Delft University of Technology, Faculty of Architecture and the Built Environment, Julianalaan 134, Delft, Netherlands

1. Introduction

Open Data in Agriculture

- definition of Open Data
- Open data in Agriculture:
 - weather data
 - seed genetics
 - environmental conditions
 - soil data
 - crop production
 - species distribution and crop data
 - and many more



1. Introduction

Dalmatian pyrethrum (*Tanacetum cinerariifolium* /Trevir./Sch. Bip.)

- plant species from Asteraceae family
- naturally distributed along the eastern Adriatic coastline
- extremely degraded habitats with shallow rocky soils (dry grasslands and garrigue)
- endemic and strictly protected plant species in Croatia
- source of pyrethrins





1. Introduction

Key stakeholders



prof. Toni Nikolić
Faculty of Science

- largest collection of botanical data in Croatia
- data from herbarium collections, field excursions, published work

- registration required for full access
- collects mostly national data
- collects only records of wild plants

1. Introduction

Key stakeholders



prof. Zlatko Šatović
Faculty of Agriculture



assoc. prof. Martina Grdiša
Faculty of Agriculture



- data on both wild and cultivated plant species
- only new and high-quality data
- completely open

- small datasets
 - records of samples included in the GeneBank

1. Introduction

The aim of this research

- examining the spatial data quality of botanical records
- source of data both CPGRD and FCD
- Dalmatian pyrethrum case study



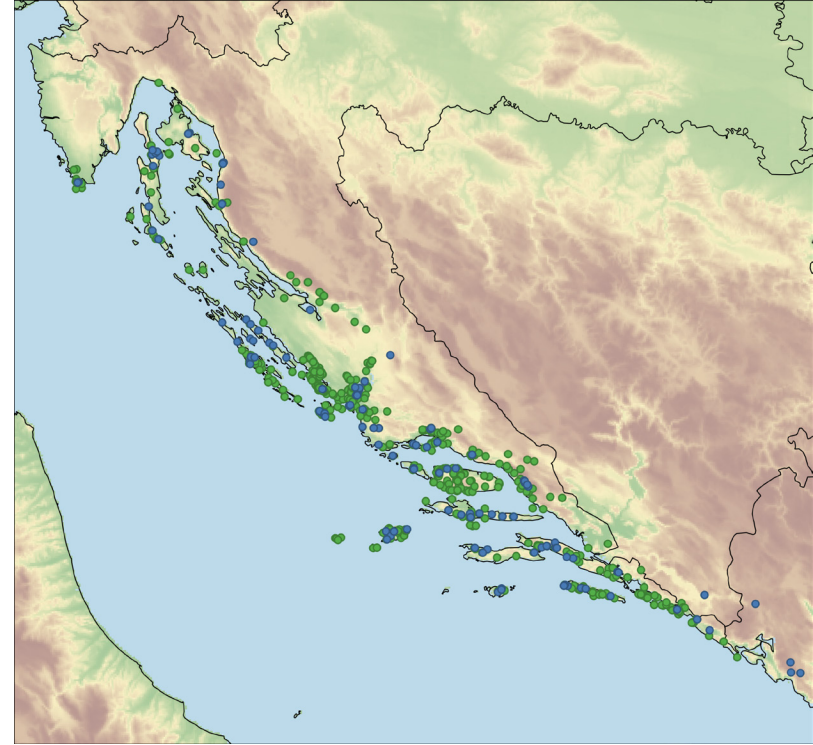
2. Materials and Methods

Data acquisition

- CPGRD - data available in open machine-readable format (.csv)
- FCD - data scrapped using BeautifulSoup version 4.9.3.(Python library)

Quality analysis

- Completeness
 - complete records (author/collector, location, collecting date, geographical coordinates)
- Logical consistency
 - consistency in data formats input for all fields (eg. DD.MM.YYYY. for collecting date field)
- Positional accuracy
 - spatial error of records (coordinates in bodies of water or outside national boundaries) in QGIS version 3.16.9.

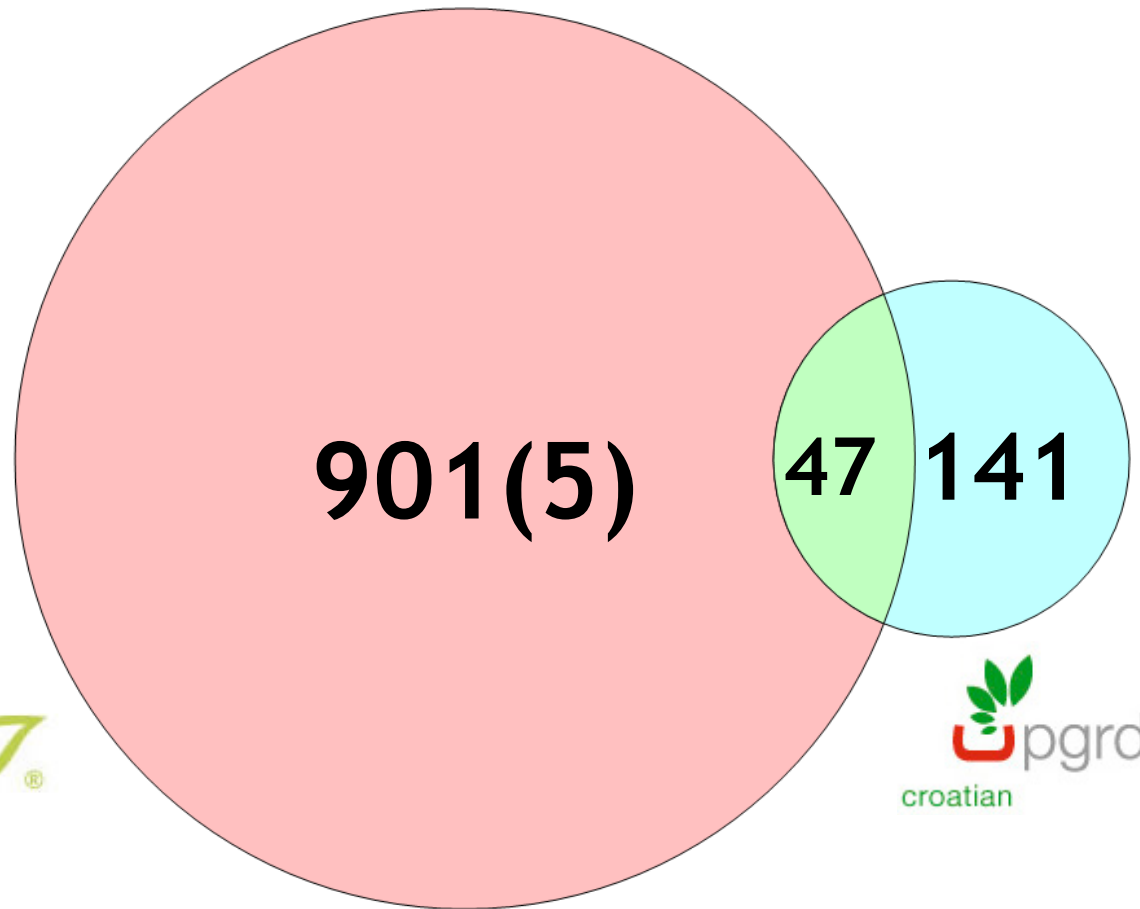




3. Results

Sample size

- in total 1047 Dalmatian pyrethrum records examined (1042)
- overlap in 47 records (1/3 of CPGRD records in FCD)





3. Results

Completeness





3. Results

Logical consistency





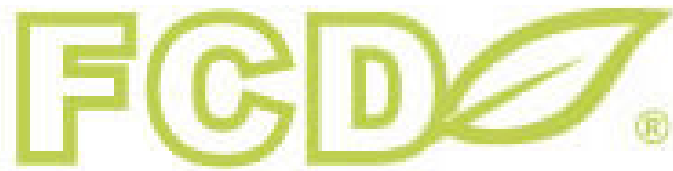
3. Results

Positional accuracy



4. Conclusions

NEW DATA INPUT
LINKING DATA



VS



DATA STRUCTURE
DATA FORMATS
TIMELINES OF THE DATA

- developing a tool for automatic assessment of positional/attribute accuracy
- expanding the quality assessment to whole databases

This research was carried out within Twinning Open Data Operational project funded within Horizon2020 Twinning programme

Thank you for your attention