12th International Conference on Information and Communication Technologies in Agriculture, Food & Environment
30 September – 3 October, 2026
Kos Island, Greece

From Scientific Literature to Decision Support: A Digital Workflow for Ecosystem Restoration Analysis

Yannis Marketakis · Michalis Mountantonakis · Pavlos Fafalios · Valantis Zervos · Anastasia Analyti · Arne Langlet · Anton Ellenbroek · Kim J. Friedman · Marcel Mallow · Yannis Tzitzikas

FORTH INSTITUTE OF COMPUTER SCIENCE
UNIVERSITY OF CRETE
TECHNICAL UNIVERSITY OF CRETE
universität wien Department of Political Science
Food and Agriculture Organization of the United Nations
THE UNIVERSITY OF WESTERN AUSTRALIA

HAICTA 2026

Outline

Motivation	3	1’
Digital Workflow	4-5	3’
Evaluation and Findings	6-8	4’
Decision Support	9	1’
Conclusions & Next steps	10	1’

Why is ecosystem restoration literature difficult to analyse?

Scientific Publications

Large & growing body of literature

Open ASFA

Google Scholar

Clarivate™ Web of Science

ProQuest

Complex Ecosystem Context

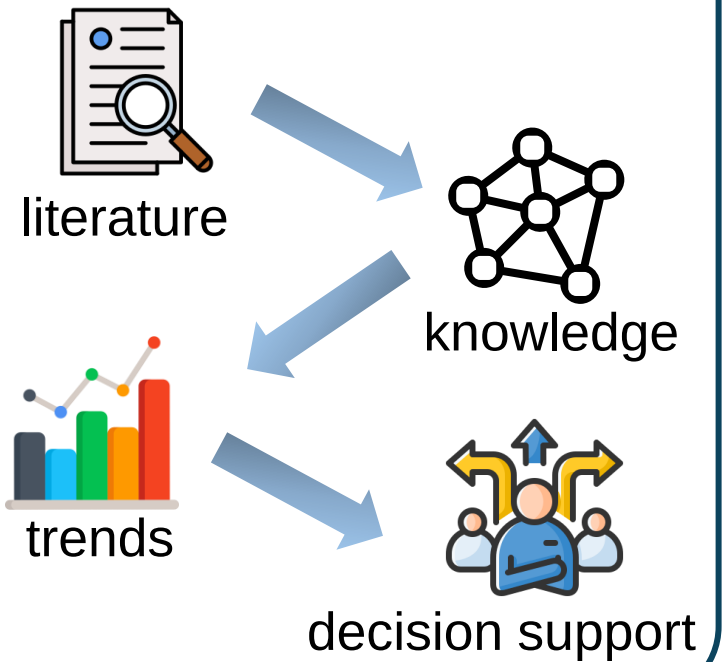
Multi-dimensional information

- Ecosystems - Terrestrial, Marine, Freshwater
- Biomes
- Species
- Habitats
- Restoration contexts

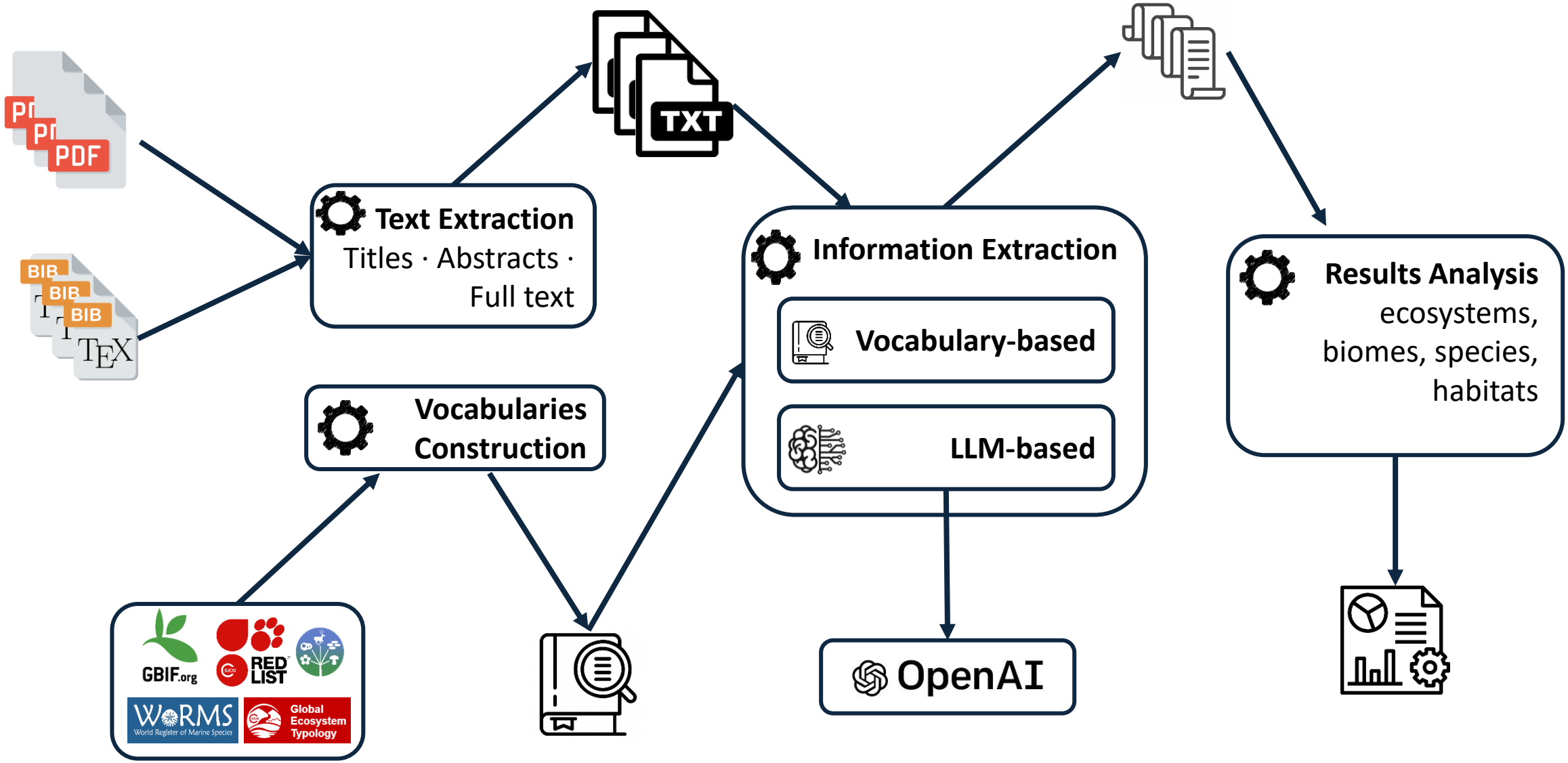
Not always explicitly mentioned in the text…

Need for Actionable Evidence

- Manual synthesis does not scale



Digital Workflow: from literature to structured knowledge



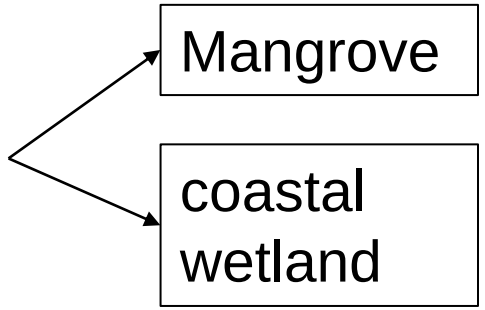
Vocabulary matching vs LLM-based classification

Vocabulary-based

Exact term matching

- Detects explicit ecosystem/species terms
- Fully transparent and reproducible
- Misses implicitly expressed context

"Mangrove restoration in coastal wetlands…"



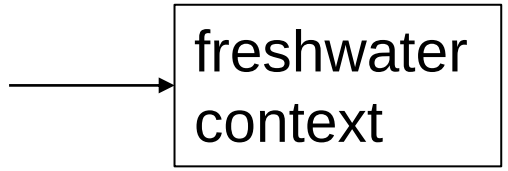
What is explicitly written?

LLM-based

Semantic interpretation

- Uses vocabularies to define target entities
- Detects explicit and implicit ecosystem context
- Goes beyond literal term matching

"restoration of spawning grounds for Atlantic salmon…"



What does the text imply?

How much text do we actually need?

Document-level classification accuracy

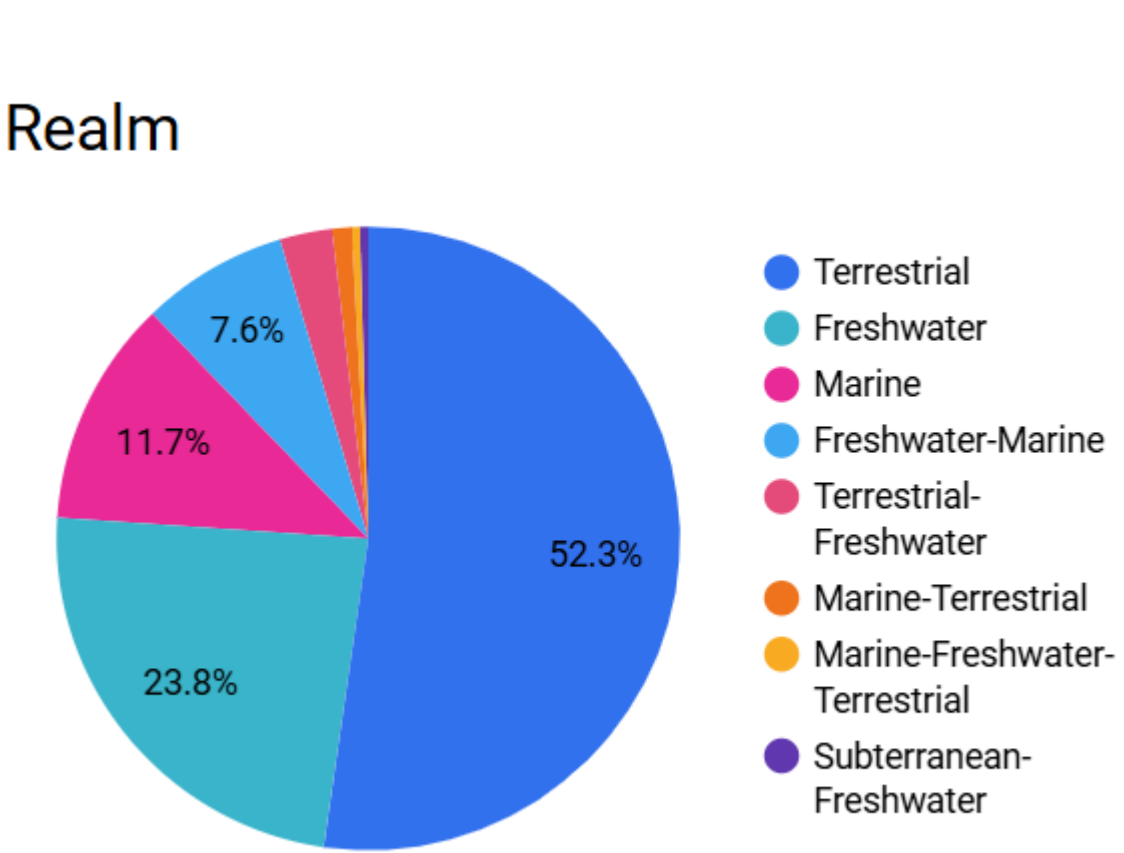
	Vocabulary-based	LLM-based
Title	5.6 %	86.7 %
Abstract	24.7 %	96.9 %
Full text	80.1 %	99.9 %

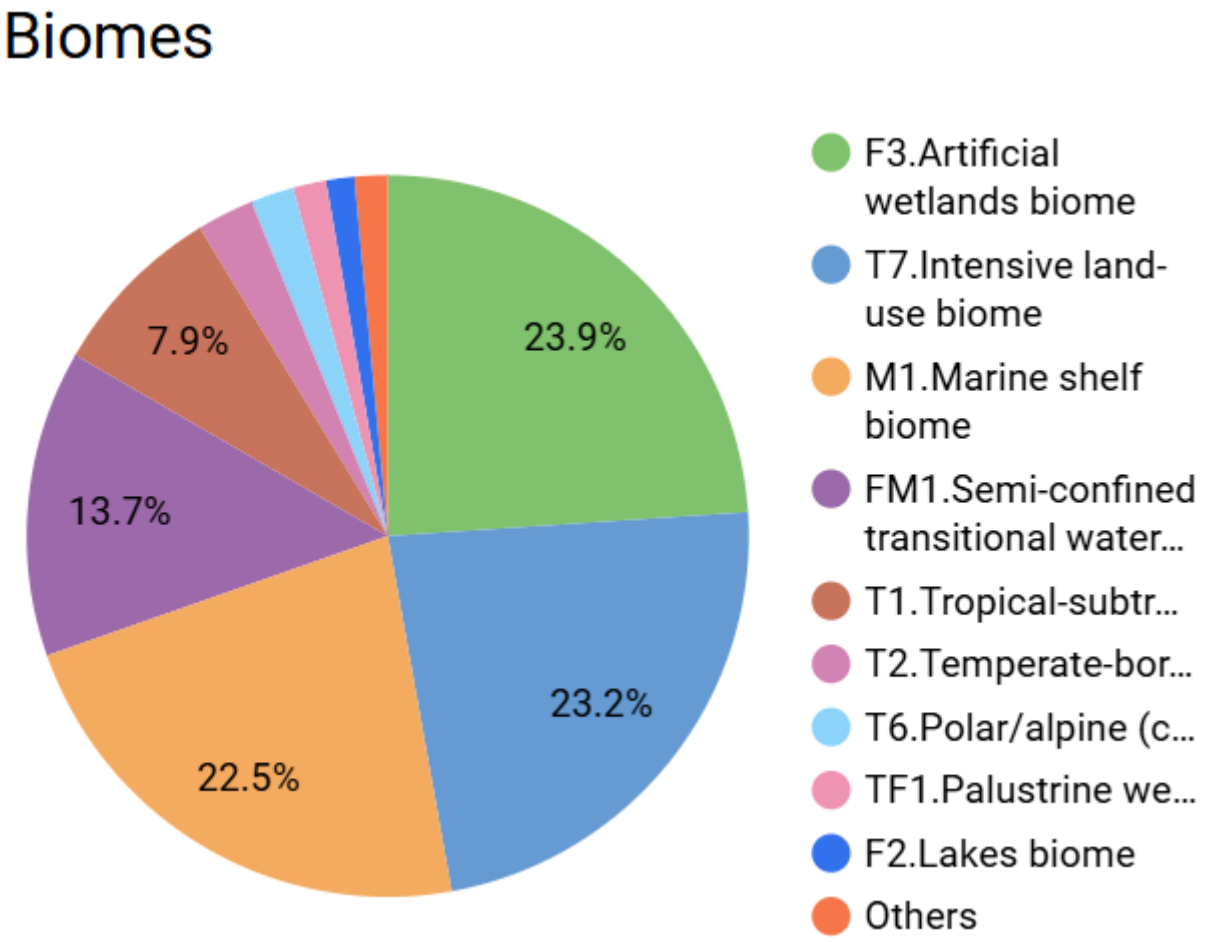
96.9 % accuracy using **only** abstracts

Near full-text performance without full-text processing

- Evaluation on two corpora
  - 639 PDF publications (full-text collection)
  - 8,776 BibTeX records (abstract-only collection)
- Model used: OpenAI GPT-5.1

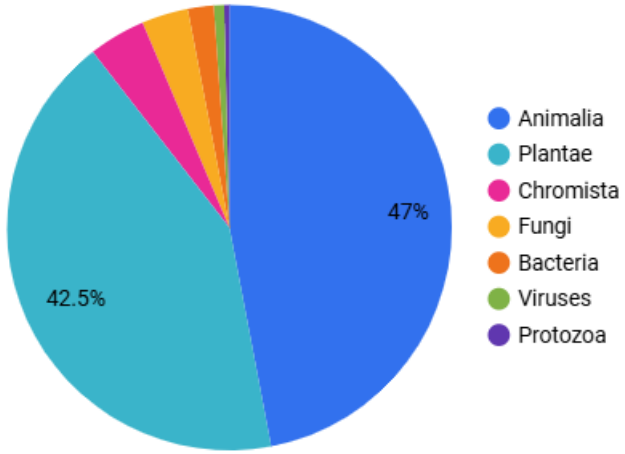
Where is restoration research concentrated?



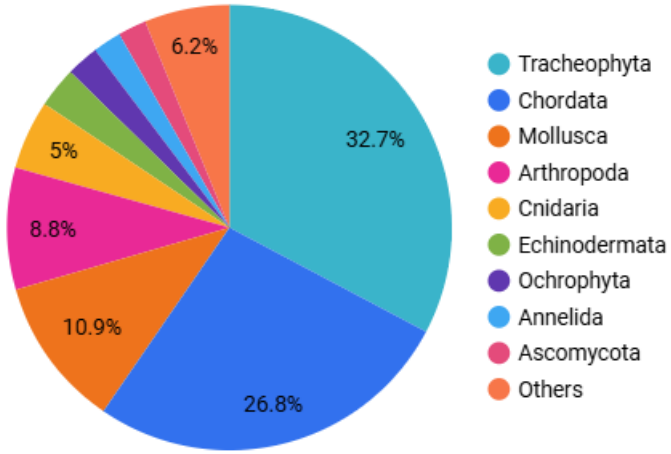


Where is restoration research concentrated? Cont'd

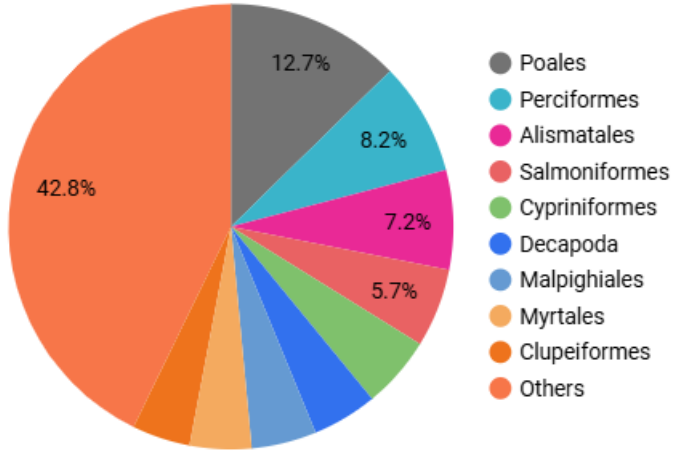
Kingdom



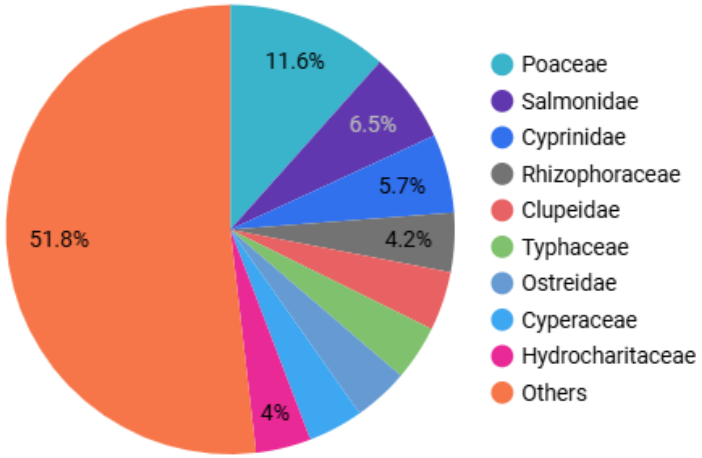
Phylum



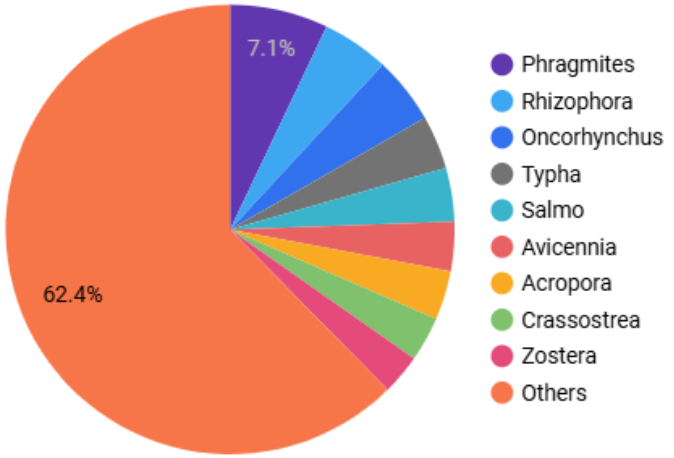
Order



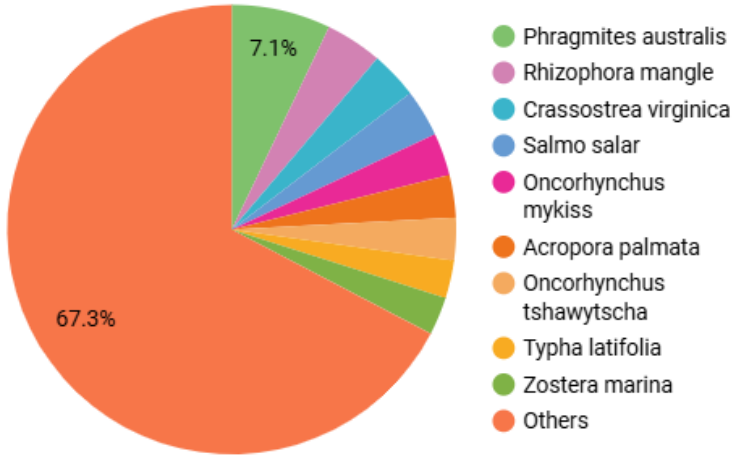
Family



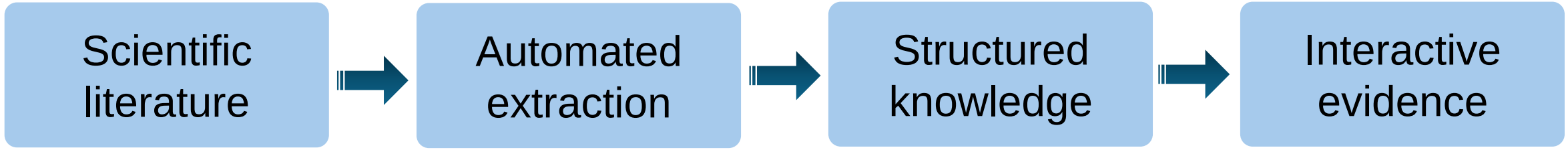
Genus



Species



From literature analysis to decision support



- **Reveal patterns** – dominance of terrestrial restoration research
- **Expose potential biases and gaps** – under-represented ecosystems or taxa become visible
- **Explore interactively** – users can filter, compare and drill down using Looker Studio
- **Produced automatically by the workflow** – the analytics are the final output of the workflow, rather than manually compiled statistics

Conclusions and Next Steps

Conclusions:

- Large-scale literature analysis can be automated
- LLM-based classification substantially improves semantic understanding
- Abstracts may be sufficient for scalable analysis

Next Steps:

- Vocabularies extension for wider coverage
- Enable question-answering over a knowledge graph from extracted information

Thank you for your attention!

Let's transform scientific literature into actionable evidence

Questions?

link to interactive analytics

Supported by :

UNITED NATIONS DECADE ON ECOSYSTEM RESTORATION 2021-2030

AIM NatuRe

FAO FIAT PANIS