Session “Beyond 2030: Building the Data Foundations for Biodiversity Action to support the KMGBF”

LIVING DATA 2025

RegRed

an effort to collate data on species' red list assessments at the national level

Florencia Grattarola, Ivo Kadlec, Adam Uličný, Kateřina Tschernosterová, Gabriel Ortega, and Petr Keil

Czech University of Life Sciences Prague
Czech Republic

Why build RegRed?

Global evaluations, like the IUCN Red List of Threatened Species™, are available as spatial data in the form of **species ranges** (>148,900 species).

Herpailurus yagouaroundi

EX EW CR EN VU NT LC

RegRed: an effort to collate data on species' red list assessments at the national level

Why build RegRed?

However, while a species may be globally safe, it can face **local extinction** or significant population decline within specific countries.

Luckily, many countries and regions also produce their own threat assessments, often compiled into **national or regional red lists**.

EX EW CR EN VU NT LC

Herpailurus yagouaroundi

RegRed: an effort to collate data on species' red list assessments at the national level

What is RegRed's goal?

Develop a comprehensive, open-access geospatial database that stores **species' threat status** at the regional, national, and subnational levels.

Phases

1. Metadata
2. Data

RegRed: an effort to collate data on species' red list assessments at the national level

1st phase: metadata

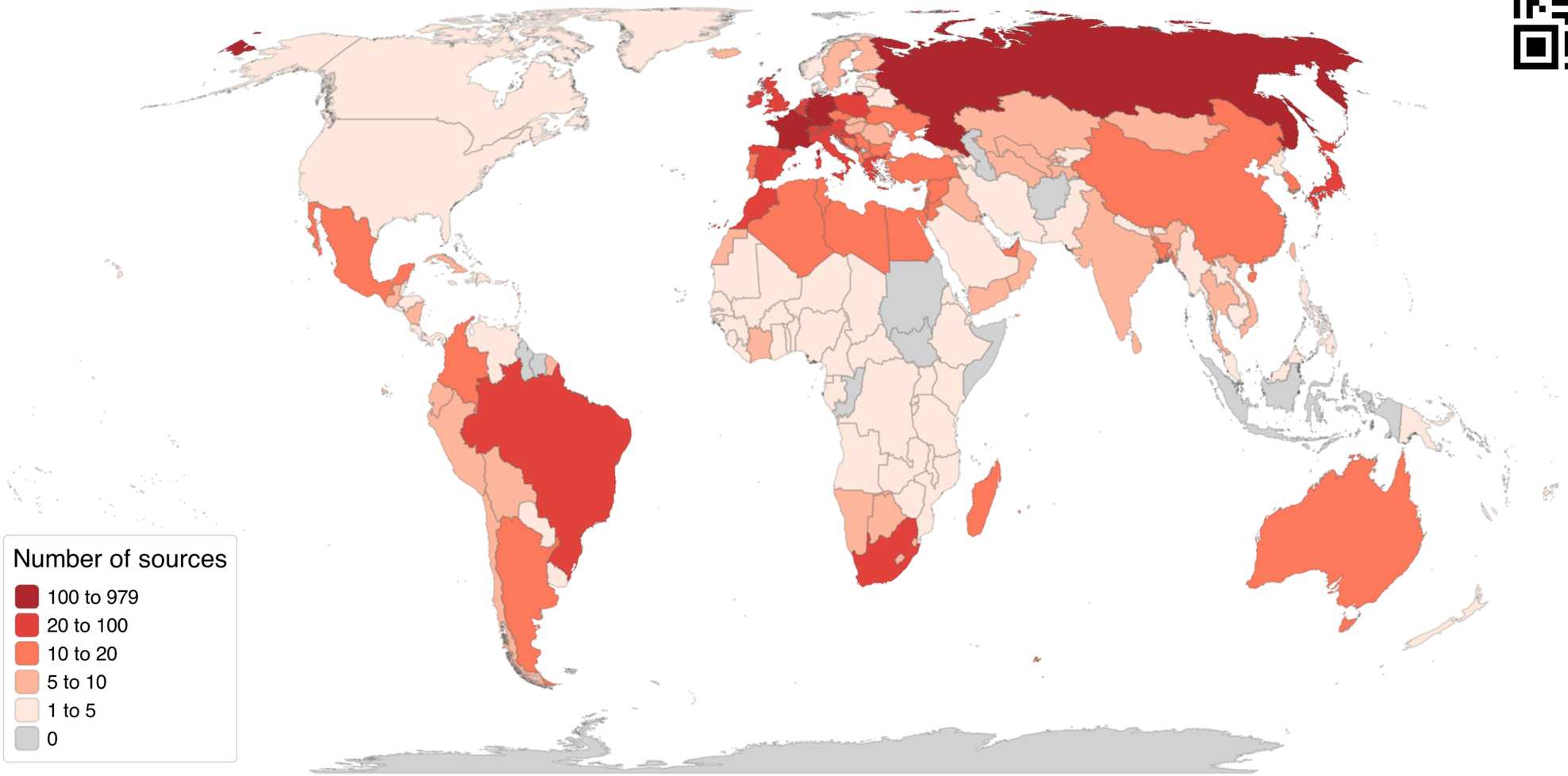
Systematic search and compilation of relevant sources worldwide.

1. Compilation of the metadata ☑
2. Publication of Data Paper ☑

https://doi.org/10.3897/arphapreprints.e160483

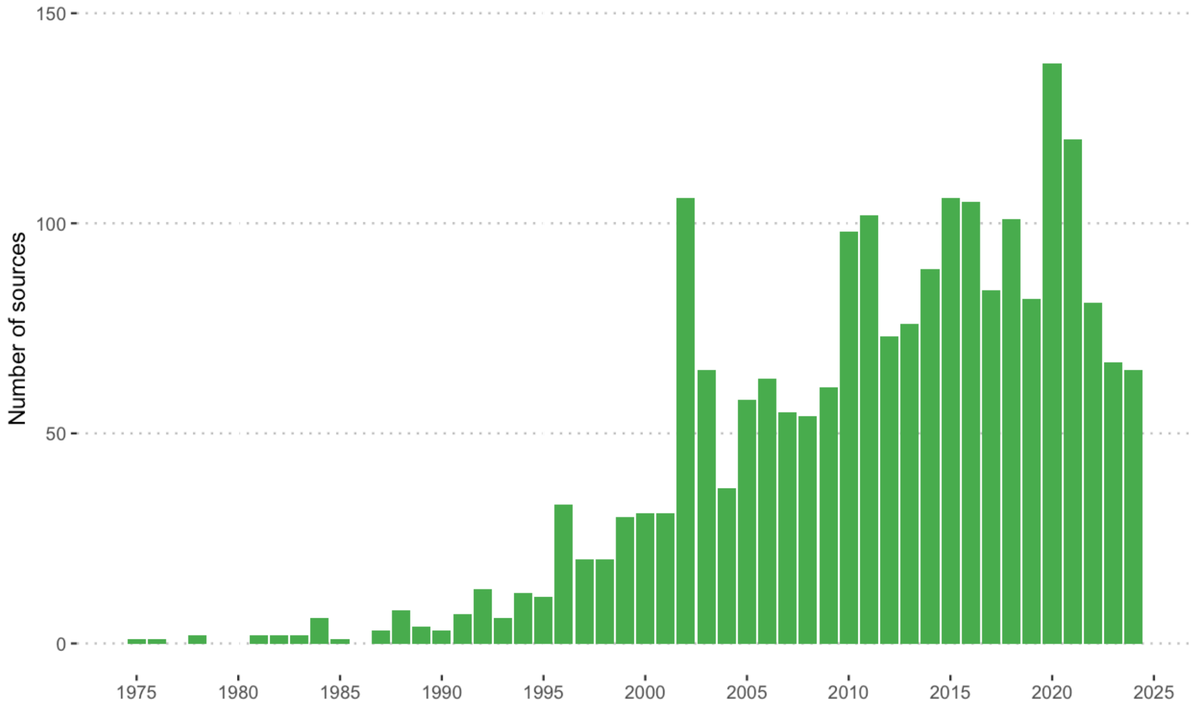
RegRed: an effort to collate data on species' red list assessments at the national level

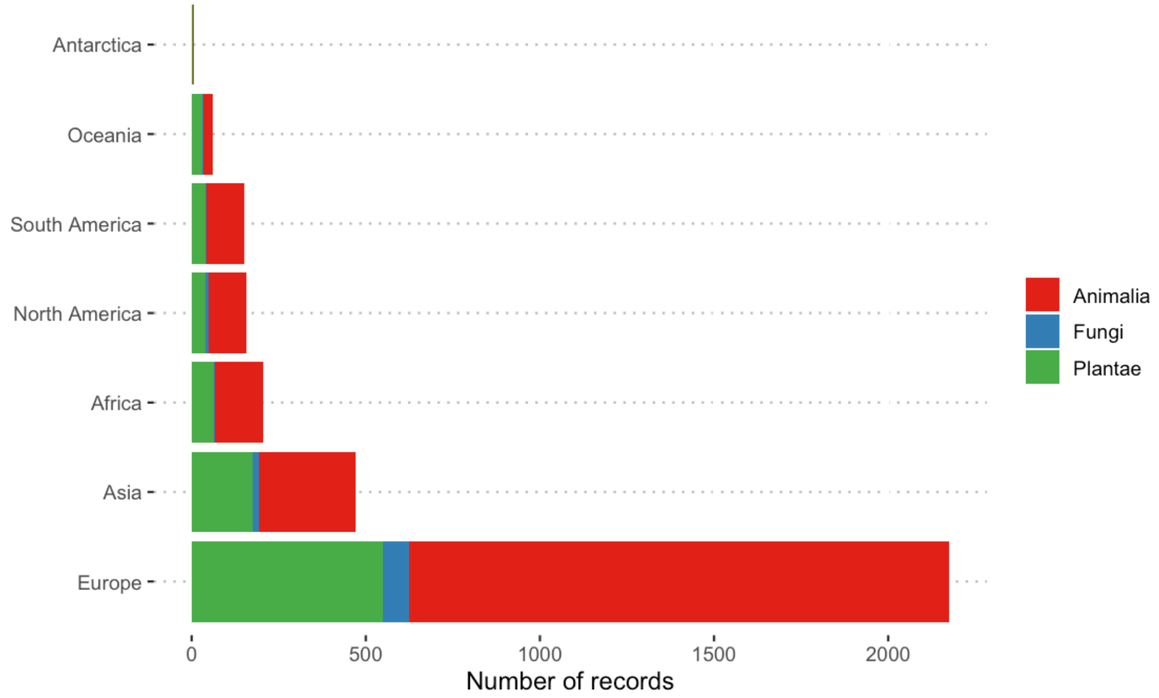
This effort yielded **2,093 sources** from **172 countries**.



RegRed: an effort to collate data on species' red list assessments at the national level

Encompassing **487 taxonomic groups**, over **50 years**.



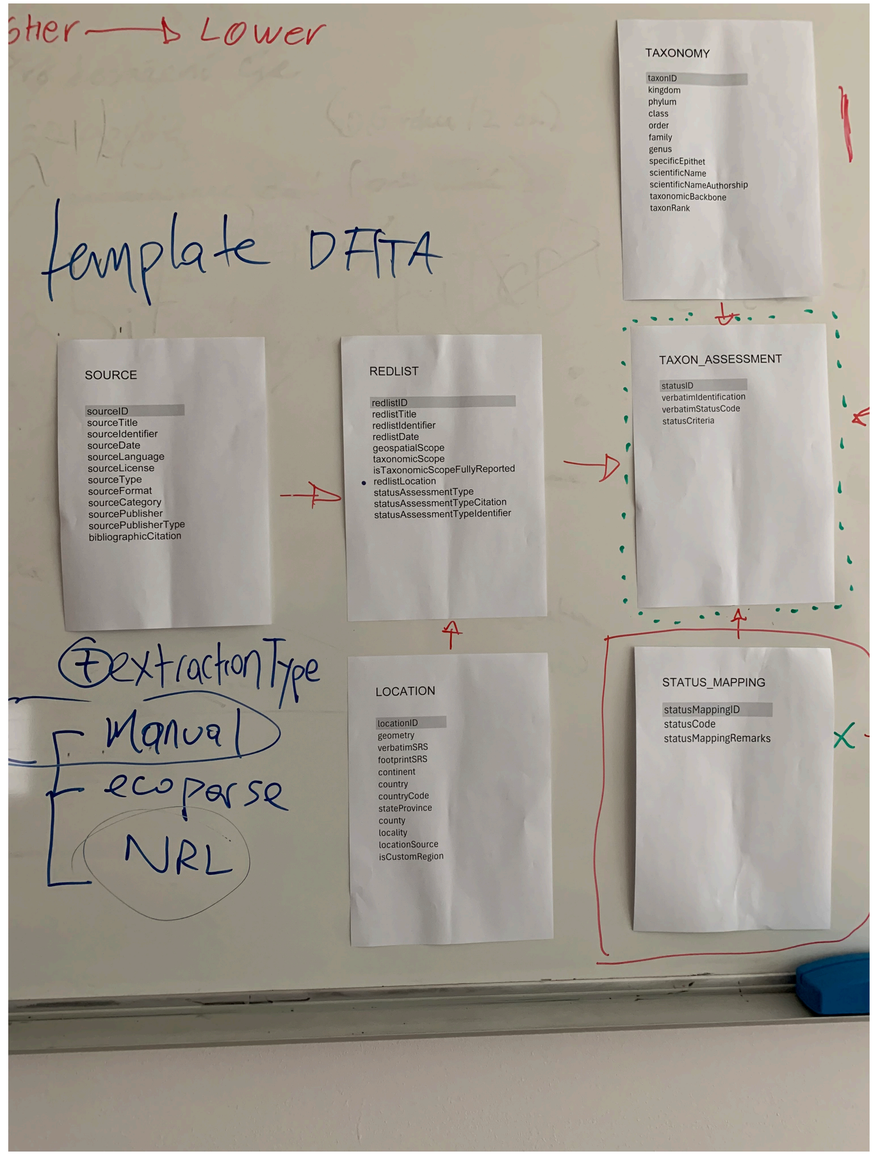


RegRed: an effort to collate data on species' red list assessments at the national level

2nd phase: data

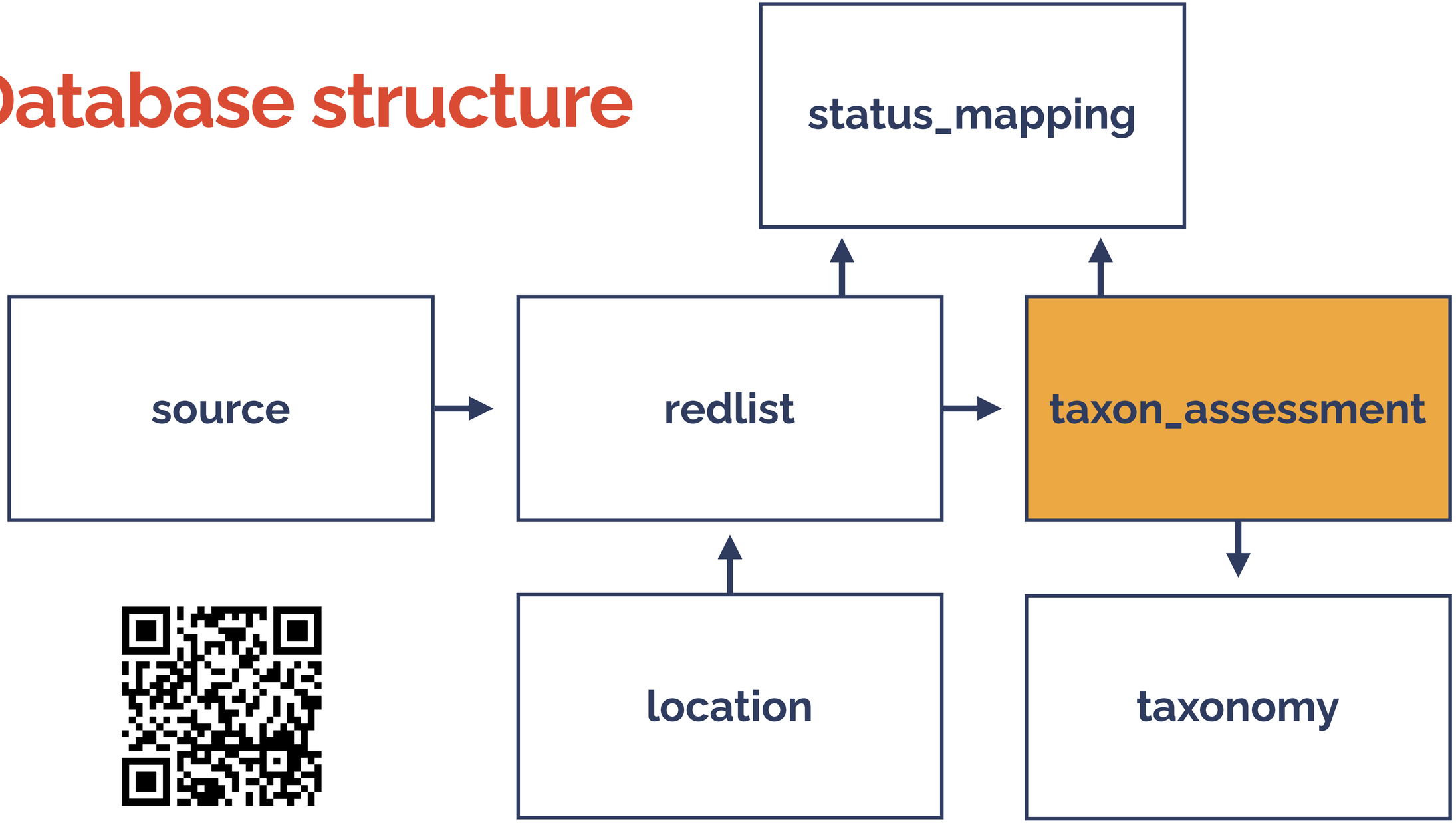
Data extraction and compilation for tetrapods.

1. Building of the database structure ☑
2. Creation of a Data Extraction Protocol ☑
3. Data extraction manually or automatically through Large Language Models (LLMs) ☐
4. Data ingestion in the PostgreSQL database ☐



RegRed: an effort to collate data on species' red list assessments at the national level

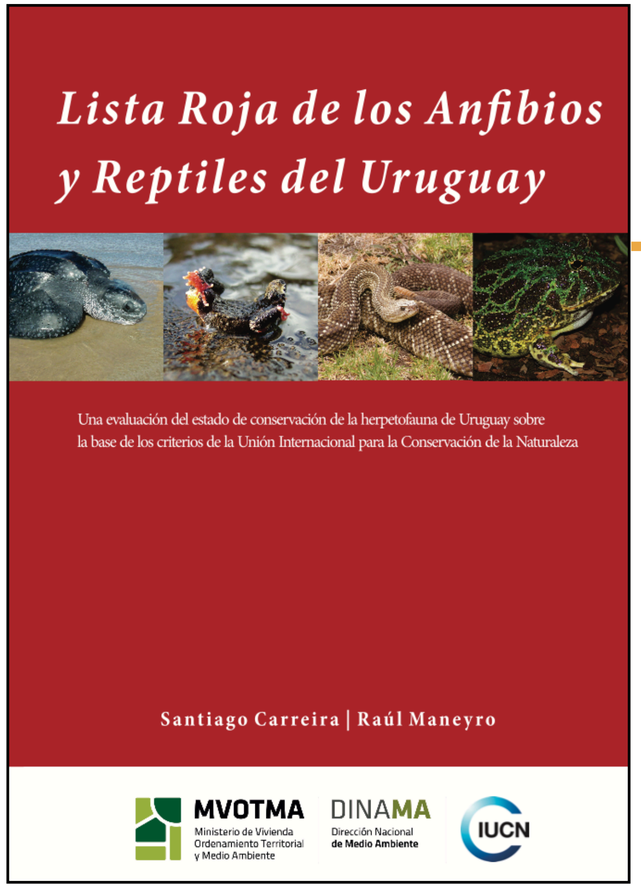
Database structure



RegRed: an effort to collate data on species' red list assessments at the national level

Example

source



status_mapping

CR | CR | Critically Endangered

redlist

Amphibians Uruguay | 2015

UICN Version 3.1

taxon_assessment

Hypsiboas albopunctatus | CR

location

South America

Uruguay

taxonomy

species | Animalia | Amphibia

Boana | albopunctata

GBIF Taxonomic backbone

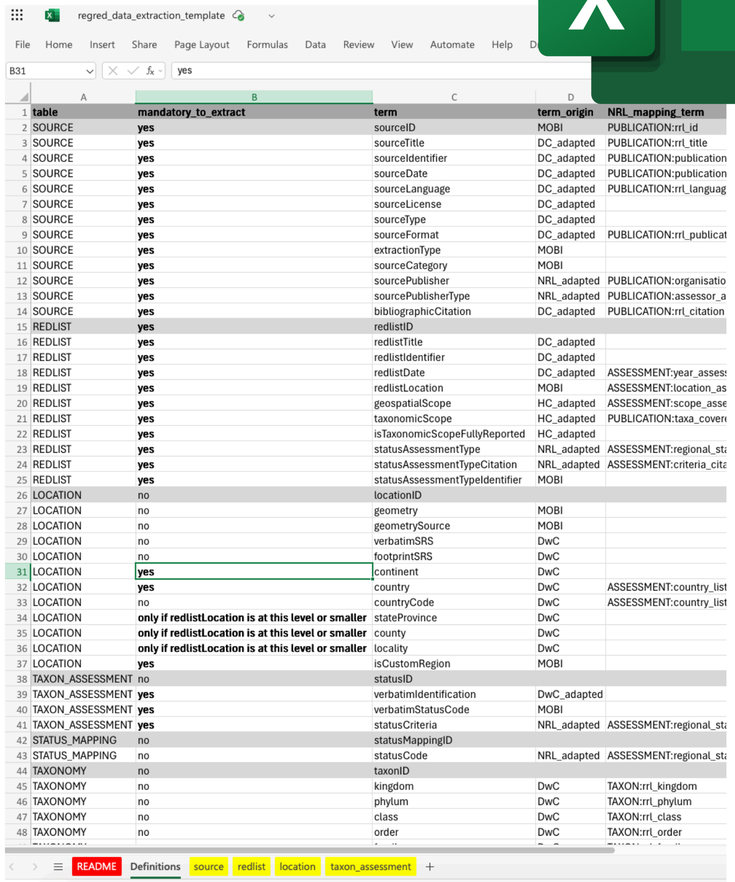
RegRed: an effort to collate data on species' red list assessments at the national level

Manual data extraction

One file per redlist

1. Fill the tables *source*, *redlist*, *location* and *taxon_assessment*
2. Map conservation status to IUCN V3.1
3. Generate bibliographic citations in RegRed's Zotero shared library
4. Harmonise taxonomy

Excel template



	A	B	C	D	E
1	**table**	**mandatory_to_extract**	**term**	**term_origin**	**NRL_mapping_term**
2	SOURCE	**yes**	sourceID	MOBI	PUBLICATION:rrl_id
3	SOURCE	**yes**	sourceTitle	DC_adapted	PUBLICATION:rrl_title
4	SOURCE	**yes**	sourceIdentifier	DC_adapted	PUBLICATION:publication
5	SOURCE	**yes**	sourceDate	DC_adapted	PUBLICATION:publication
6	SOURCE	**yes**	sourceLanguage	DC_adapted	PUBLICATION:rrl_languag
7	SOURCE	**yes**	sourceLicense	DC_adapted	
8	SOURCE	**yes**	sourceType	DC_adapted	
9	SOURCE	**yes**	sourceFormat	DC_adapted	PUBLICATION:rrl_publicat
10	SOURCE	**yes**	extractionType	MOBI	
11	SOURCE	**yes**	sourceCategory	MOBI	
12	SOURCE	**yes**	sourcePublisher	NRL_adapted	PUBLICATION:organisatio
13	SOURCE	**yes**	sourcePublisherType	NRL_adapted	PUBLICATION:assessor_a
14	SOURCE	**yes**	bibliographicCitation	DC_adapted	PUBLICATION:rrl_citation
15	REDLIST	**yes**	redlistID		
16	REDLIST	**yes**	redlistTitle	DC_adapted	
17	REDLIST	**yes**	redlistIdentifier	DC_adapted	
18	REDLIST	**yes**	redlistDate	DC_adapted	ASSESSMENT:year_assess
19	REDLIST	**yes**	redlistLocation	MOBI	ASSESSMENT:location_as
20	REDLIST	**yes**	geospatialScope	HC_adapted	ASSESSMENT:scope_asse
21	REDLIST	**yes**	taxonomicScope	HC_adapted	PUBLICATION:taxa_cover
22	REDLIST	**yes**	isTaxonomicScopeFullyReported	HC_adapted	
23	REDLIST	**yes**	statusAssessmentType	NRL_adapted	ASSESSMENT:regional_st
24	REDLIST	**yes**	statusAssessmentTypeCitation	NRL_adapted	ASSESSMENT:criteria_cita
25	REDLIST	**yes**	statusAssessmentTypeIdentifier	MOBI	
26	LOCATION	no	locationID		
27	LOCATION	no	geometry	MOBI	
28	LOCATION	no	geometrySource	MOBI	
29	LOCATION	no	verbatimSRS	DwC	
30	LOCATION	no	footprintSRS	DwC	
31	LOCATION	**yes**	continent	DwC	
32	LOCATION	**yes**	country	DwC	ASSESSMENT:country_list
33	LOCATION	no	countryCode	DwC	ASSESSMENT:country_list
34	LOCATION	**only if redlistLocation is at this level or smaller**	stateProvince	DwC	
35	LOCATION	**only if redlistLocation is at this level or smaller**	county	DwC	
36	LOCATION	**only if redlistLocation is at this level or smaller**	locality	DwC	
37	LOCATION	**yes**	isCustomRegion	MOBI	
38	TAXON_ASSESSMENT	no	statusID		
39	TAXON_ASSESSMENT	**yes**	verbatimIdentification	DwC_adapted	
40	TAXON_ASSESSMENT	**yes**	verbatimStatusCode	MOBI	
41	TAXON_ASSESSMENT	**yes**	statusCriteria	NRL_adapted	ASSESSMENT:regional_st
42	STATUS_MAPPING	no	statusMappingID		
43	STATUS_MAPPING	no	statusCode	NRL_adapted	ASSESSMENT:regional_st
44	TAXONOMY	no	taxonID		
45	TAXONOMY	no	kingdom	DwC	TAXON:rrl_kingdom
46	TAXONOMY	no	phylum	DwC	TAXON:rrl_phylum
47	TAXONOMY	no	class	DwC	TAXON:rrl_class
48	TAXONOMY	no	order	DwC	TAXON:rrl_order

RegRed: an effort to collate data on species' red list assessments at the national level

Automated data extraction (using LLMs)

Check Adam Uličný's poster!

https://www.livingdata2025.com/posters.html?poster=7011700

Extracting Species-Level Threat Data from Heterogeneous Conservation Literature: Towards a Scalable Approach Using LLMs

Copy ID | Share

Authors

Adam Ulicny - Czech University of Life Sciences, Czechia

Ivo Kadlec - Czech University of Life Sciences, Czechia

Petr Keil - Czech University of Life Sciences, Czechia

Florencia Grattarola - Czech University of Life Sciences, Czechia

EN | REGULAR | 7011700

Abstract

Large language models (LLMs) are increasingly being used to automate knowledge extraction from unstructured text. We explore their large scale application in parsing species threat assessments from conservation literature, such as national Red Lists. Conservation status data is often buried in lengthy, unstructured PDFs, making it time-consuming to track and analyze species-level assessments. Despite advances in natural language processing (NLP), few tools can extract structured data across diverse document formats with sufficient accuracy. We present a concept of an end-to-end pipeline that tests a combination of techniques, including species name recognition, contextual chunking, and both local and remote LLM APIs (e.g., Gemini, Ollama) to extract relevant data. Outputs are subject to automated checks and manual validation to ensure quality. Preliminary results show us species name and threat level data from a 600 page PDF can be extracted in around 30 seconds. This approach balances automation with human oversight, enabling efficient and scalable parsing of documents. The pipeline serves as a practical framework for researchers seeking to extract various kinds of species-level data from existing literature.

RegRed: an effort to collate data on species' red list assessments at the national level

In summary, RegRed's cool features

1. Geospatial relational database.
2. Highly flexible: taxonomy, locations, and threatened statuses.
3. Borrows terms from the DwC and HC standards.
4. Compatible with NRL (the National Red List project).
5. Easy data extraction workflow for non-experts, manual and automated.

RegRed: an effort to collate data on species' red list assessments at the national level

Who's part of this effort?

Leading the project

tetrapods

Florencia Grattarola
Petr Keil
Ivo Kadlec
Adam Uličný
Kateřina Tschernosterová
Gabriel Ortega
Safiya Praleskouskaya

Partners

plants

Laura Méndez
Alexander Zizka
Marten Winter
Sonja Knapp
Erik Welk

insects

Shawan Chowdhury

fish

Aurele Toussaint

Sophie Ledger
Neil Cox

National Red List

RegRed: an effort to collate data on species' red list assessments at the national level

How can you engage?

Get in contact!

RegRed@fzp.czu.cz

Follow our project on GitHub

https://github.com/RegRed-project

RegRed: an effort to collate data on species' red list assessments at the national level

¡Gracias! Thanks!

Attribution
Status_iucn3.1.svg - Peng Halasz (CC-BY via Wikimedia Commons)
Herpailurus yagouaroundi - Gabriela Palomo-Muñoz (CC-BY-NC via Phylopic)
Icons from Freepik via Flaticon.com
All other figures were created by Florencia Grattarola

LIVING DATA 2025

Czech University of Life Sciences Prague

MOBI Lab

BEAST Project

Funded by the European Union

European Research Council
Established by the European Commission

This work was funded by the European Union (ERC, BEAST, 101044740). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

flograttarola.com | @flograttarola.bsky.social | @flograttarola | ecoevo.social/@flograttarola