

Checklistbank, use of different taxonomies and Catalogue of Life

Dave Martin



Agenda

- ChecklistBank
- matching-ws
- Multi taxonomy support in GBIF API
- Processing as service





ChecklistBank

Taxonomic data infrastructure for everyone

The screenshot shows the ChecklistBank homepage. At the top is a banner with a close-up image of orange and yellow flowers. Overlaid on the banner is the ChecklistBank logo and the text "Index and repository for taxonomic data". Below the banner are four summary statistics:

Species in Catalogue of Life	Name Usages in ChecklistBank	Datasets in ChecklistBank	Latest COL Checklist
2,177,735	350,244,666	53,447	2024-07-18

Below these stats is a section about the Catalogue of Life (COL). It explains COL's purpose, its collaboration with GBIF, and how taxonomic communities can publish checklists. It also mentions the ChecklistBank API and login requirements.

At the bottom of the page, there's footer information including developer credits, feedback links, and version numbers.

Open data, open access
taxonomic data publishing
platform

Tools for:

1. Importing lists
2. Comparing lists
3. Building lists

checklistbank.org

api.checklistbank.org



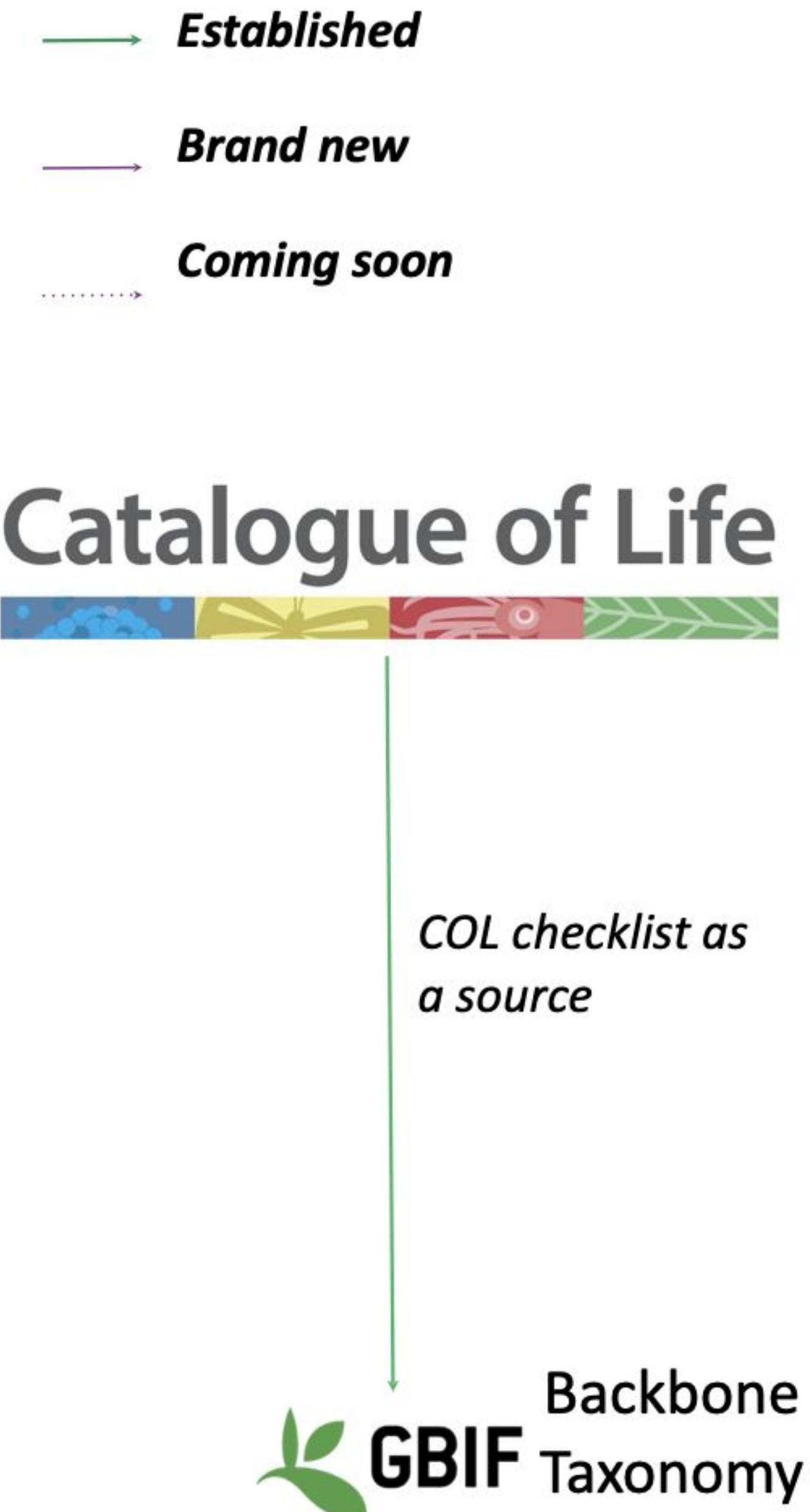
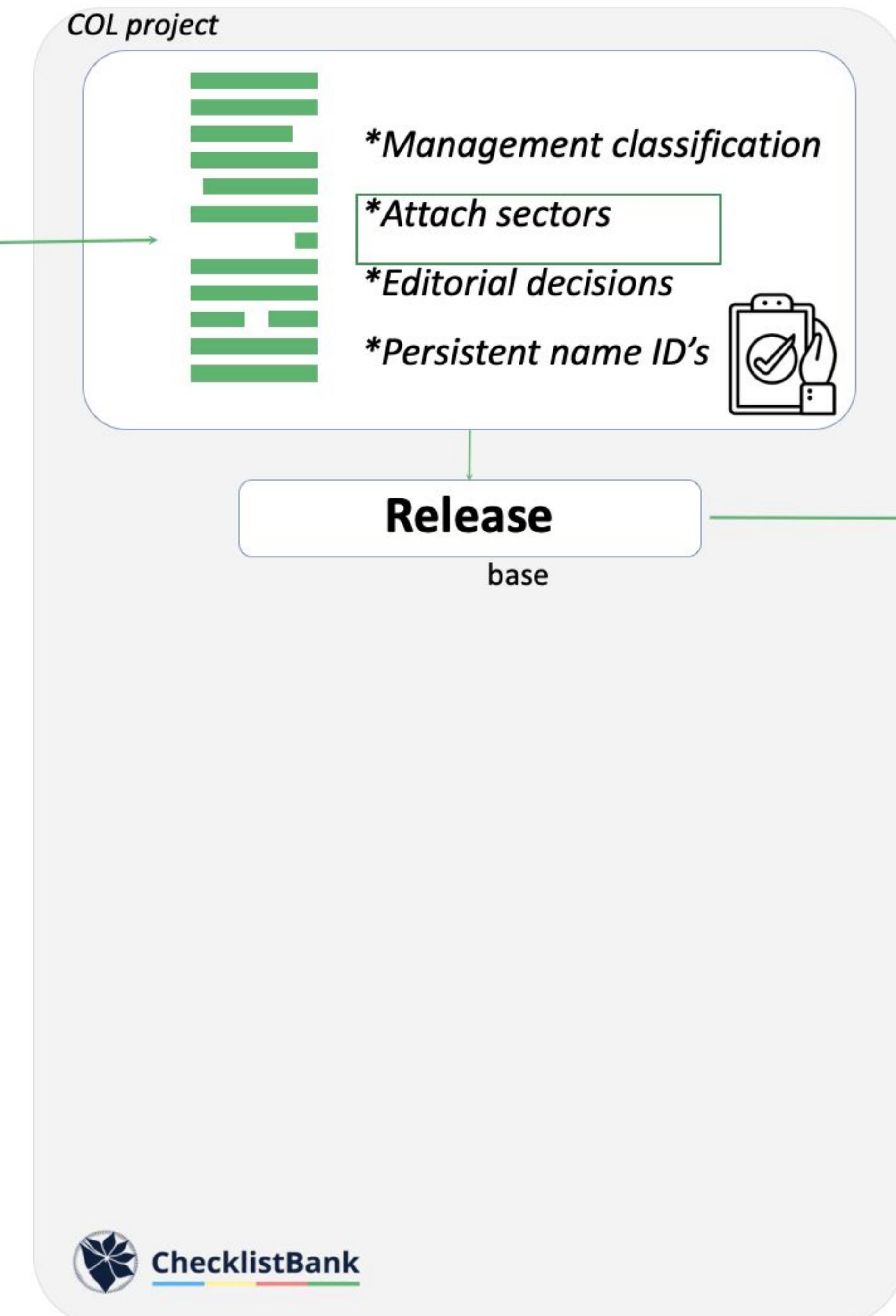
Catalogue of Life

Global checklists



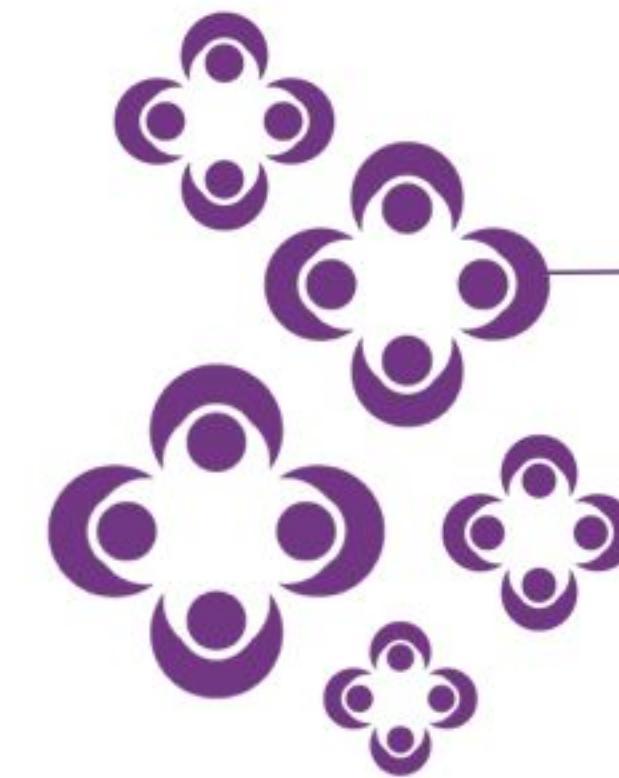
Taxonomic Communities

Regional / management checklists

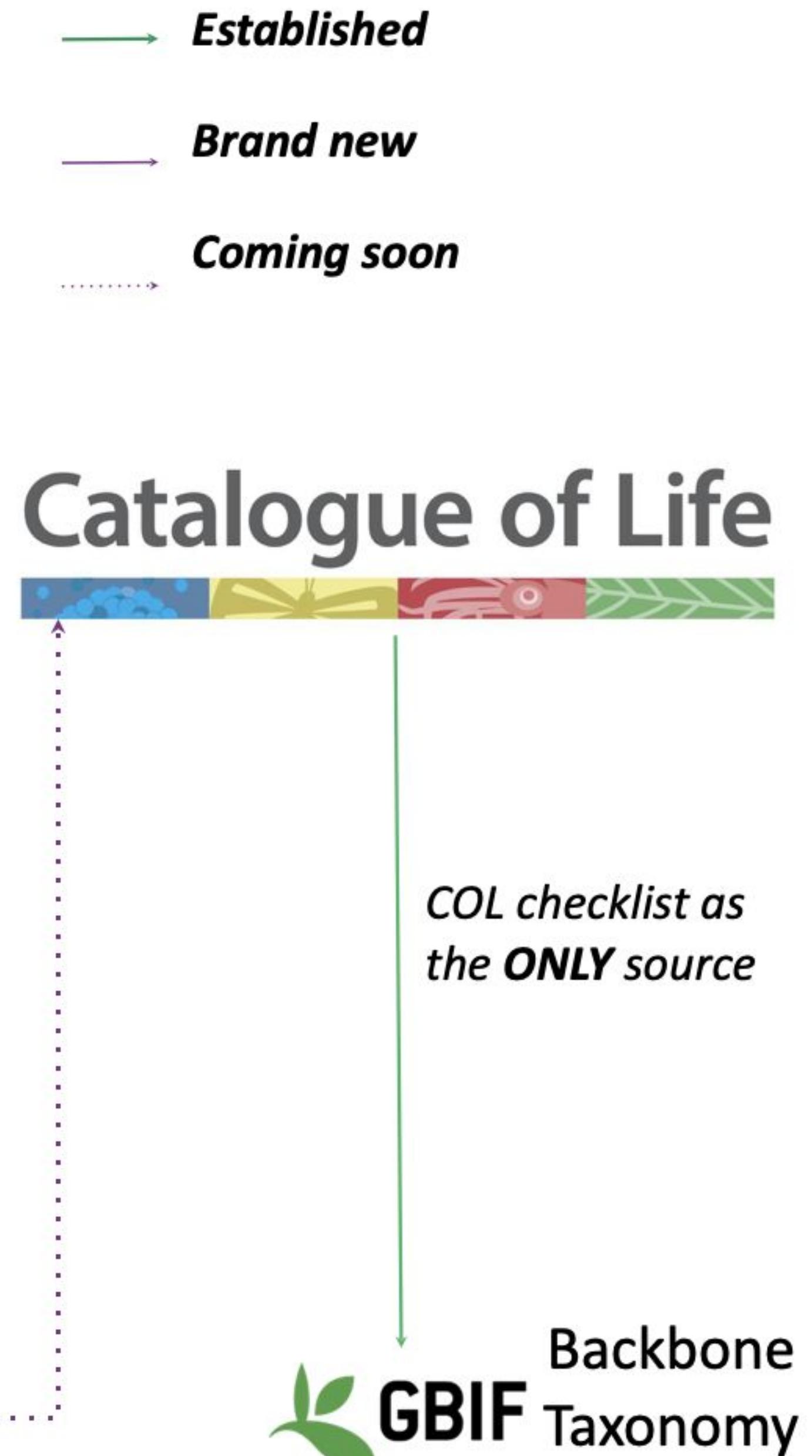
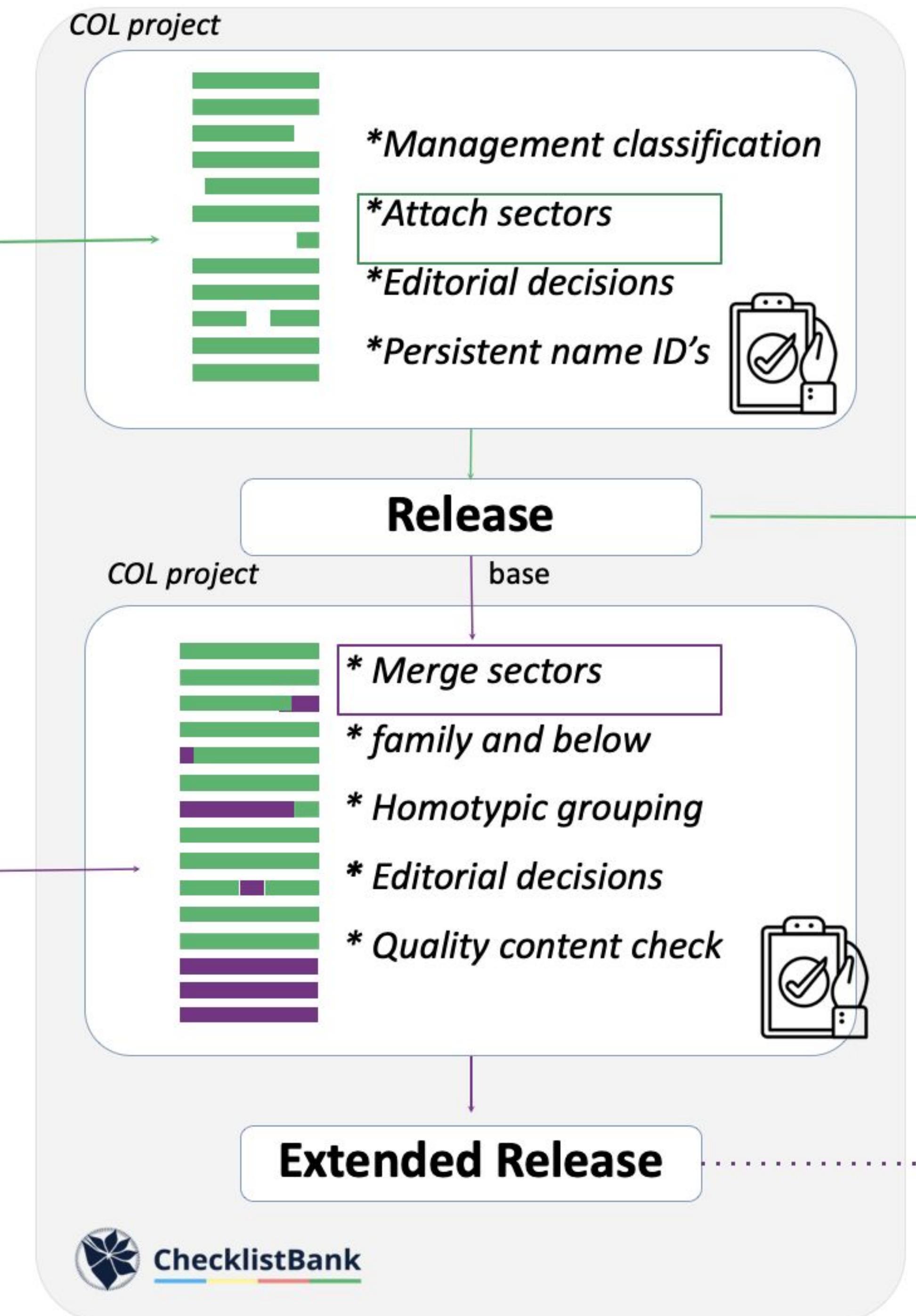


Global checklists

Regional / management checklists



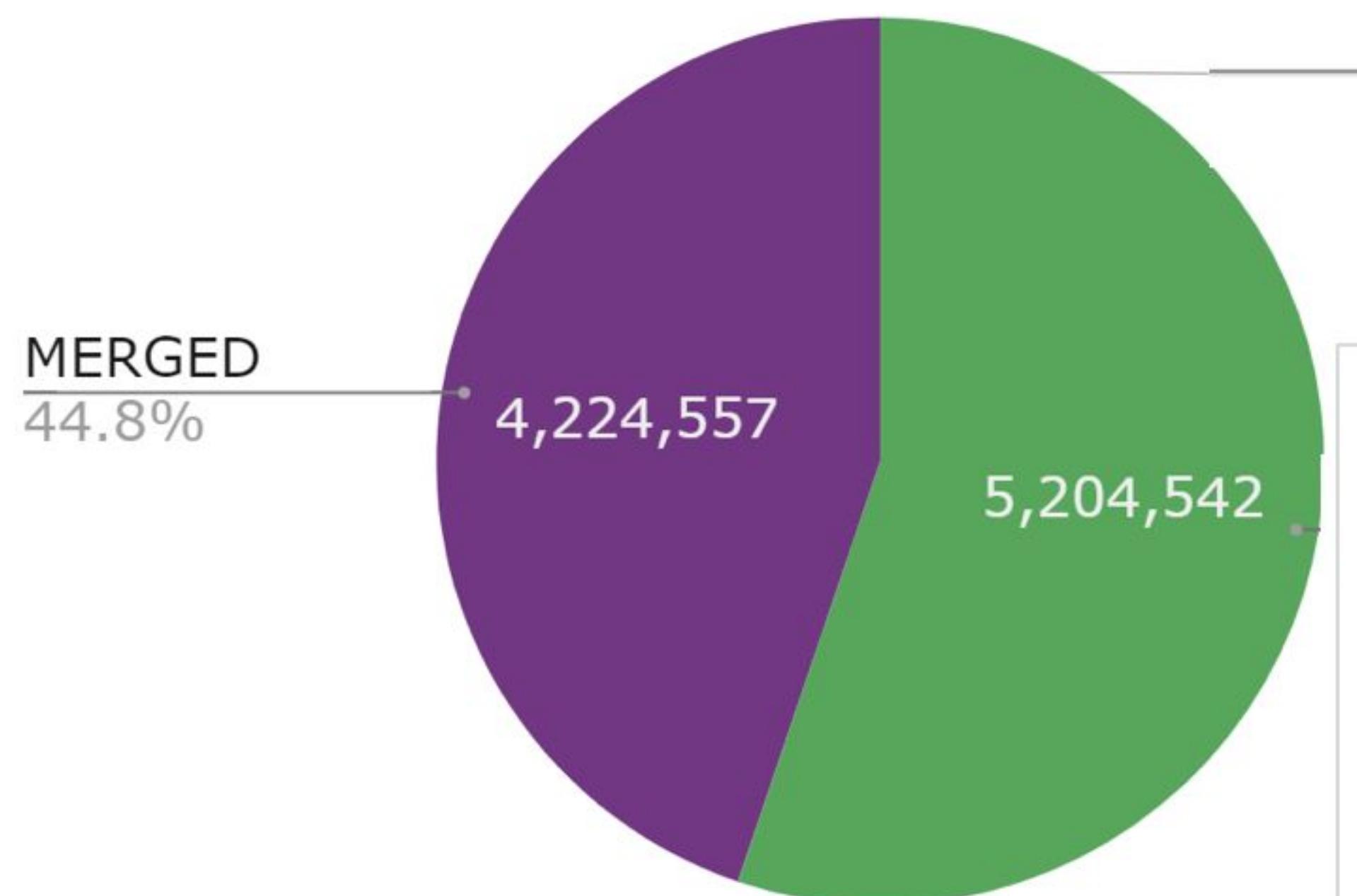
Taxonomic Communities



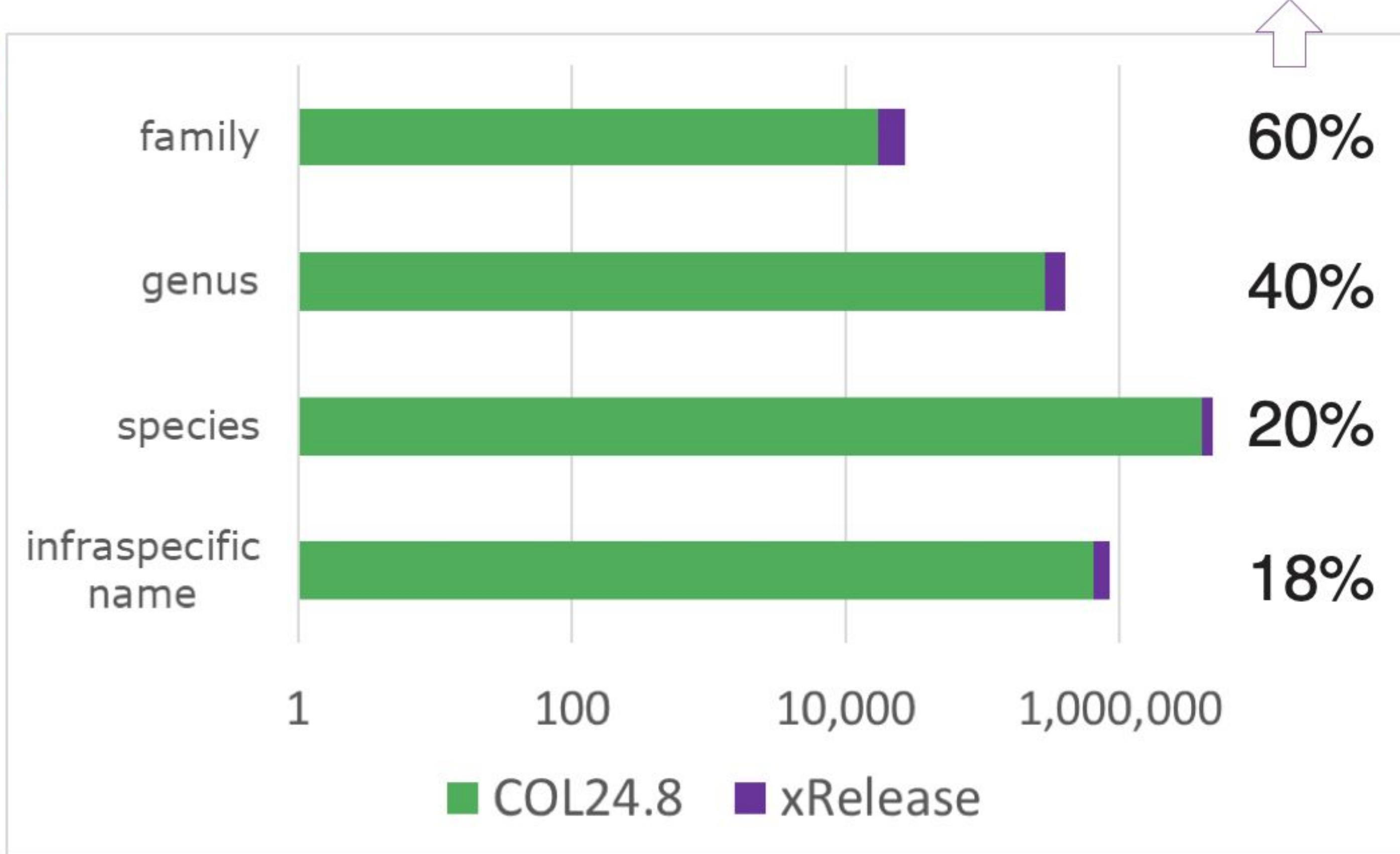
Merged names in COL (all ranks)

xRelease
2024-08-25

Name count

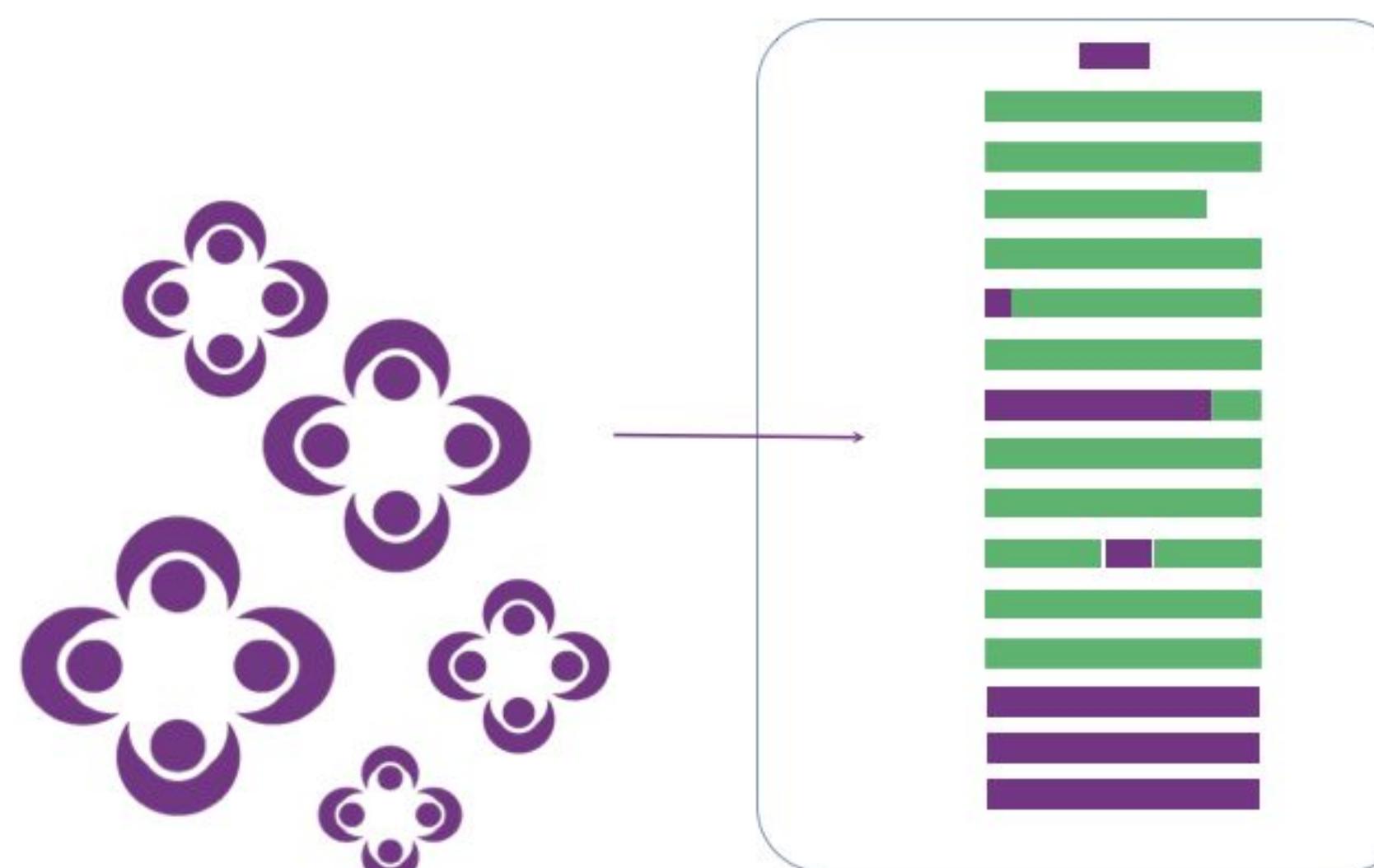


*ONLY family names and
below are being merged*



You can contribute to a better COL

1. Publish or recommend a checklist



2. Issues report

The screenshot shows a GitHub issue page with the title "Missing Chrysididae species #622". The issue was opened by mdoering last month and has 3 comments. A comment from mdoering states: "There are quite a few species missing from the [Chrysididae](#) family in COL 24.1. The following species were existing in the SANBI species dictionary, but not found in COL:" followed by a bulleted list of species names.

Missing Chrysididae species #622

Open mdoering opened this issue last month · 3 comments

mdoering commented last month

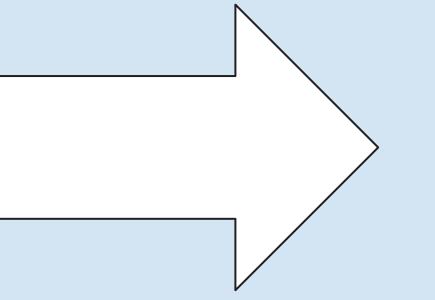
There are quite a few species missing from the [Chrysididae](#) family in COL 24.1. The following species were existing in the SANBI species dictionary, but not found in COL:

- *Acrotoma arnoldi*
- *Allococelia bidens*
- *Allococelia glabra*
- *Allococelia latinota*
- *Allococelia mocsaryi*
- *Brugmoia torrida*
- *Chrysidea minima*
- *Chrysidea pumila*
- *Chrysis alecto*
- *Chrysis alternans*
- *Chrysis angustula*

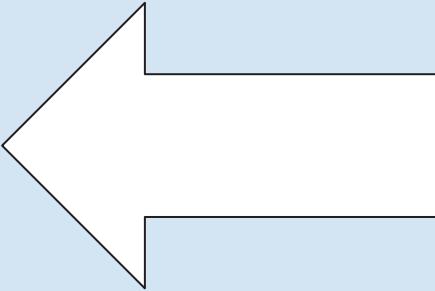
<https://github.com/CatalogueOfLife/data/issues>

Matching-ws

ChecklistBank



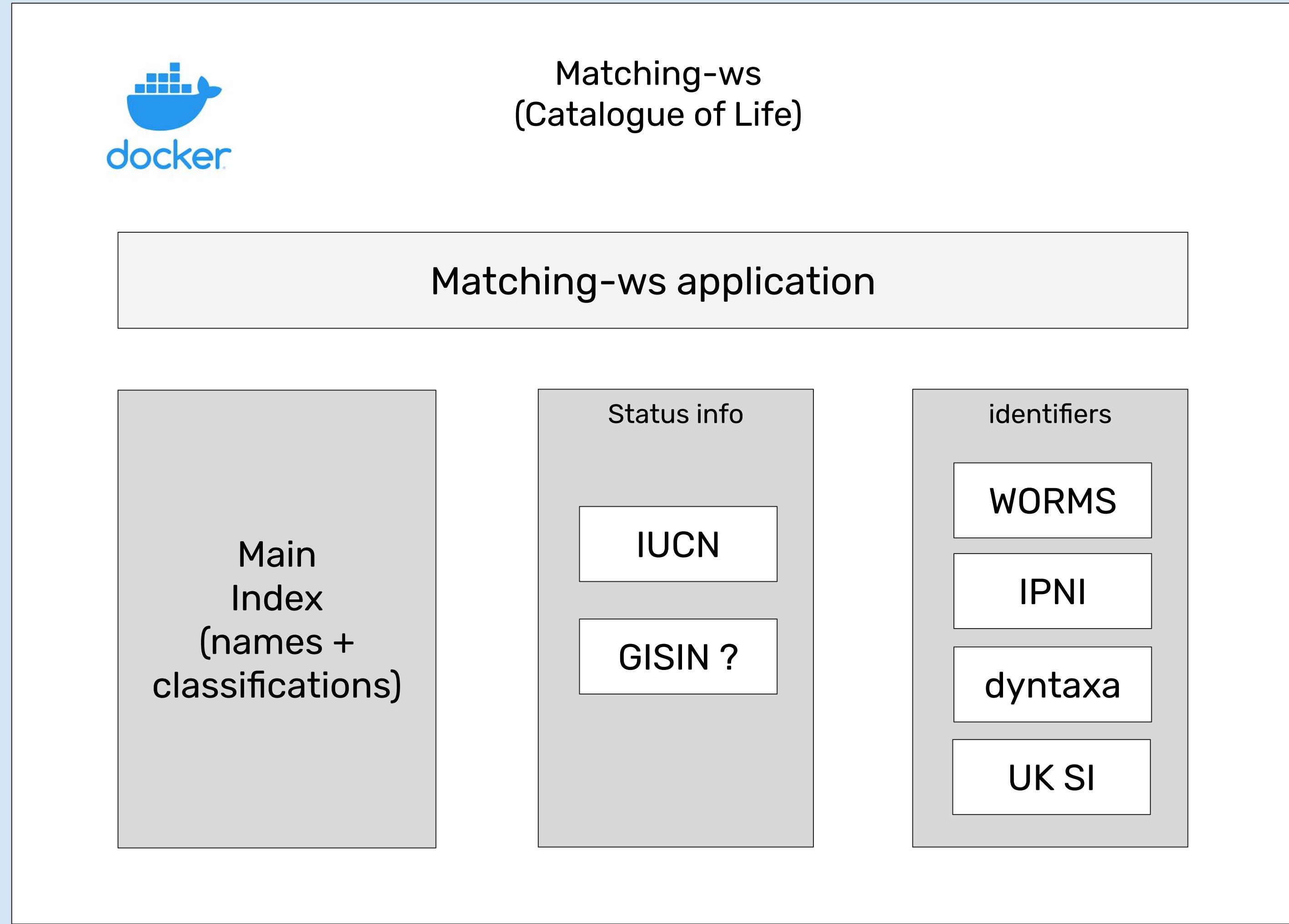
matching-ws



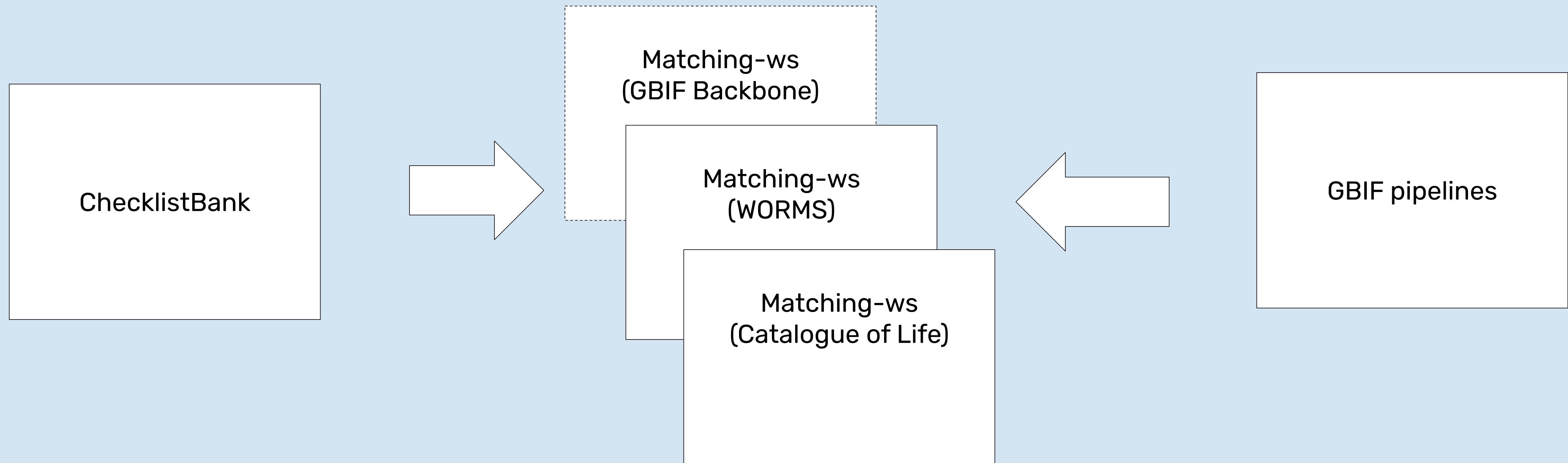
GBIF pipelines



Matching-ws



Matching-ws



<https://repository.gbif.org>

The screenshot shows the Sonatype Nexus Repository interface at <https://repository.gbif.org/>. The left sidebar has a 'Docker' tab selected. The main area is a search interface for Docker components, with fields for 'Image Name' (set to 'matching-ws'), 'Image Tag' (set to 'Any'), 'Layer Id' (set to 'Any'), and 'Content Digest' (set to 'Any'). A 'More criteria' button is available. Below the search form is a table listing Docker images. The table columns are: Name, Last Updated, Version, Format, Repository, and a link icon. There are approximately 20 entries for 'matching-ws' images, mostly from 'xcol-latest' and 'worms-latest' repositories.

Name	Last Updated	Version	Format	Repository	
matching-ws	12/03/2025, 01:01:33	xcol-latest-arm64	docker	docker	>
matching-ws	12/03/2025, 00:56:26	xcol-latest-amd64	docker	docker	>
matching-ws	12/03/2025, 00:56:28	xcol-latest	docker	docker	>
matching-ws	10/03/2025, 22:27:59	worms-latest-arm64	docker	docker	>
matching-ws	10/03/2025, 22:27:04	worms-latest-amd64	docker	docker	>
matching-ws	10/03/2025, 22:27:05	worms-latest	docker	docker	>
matching-ws	10/03/2025, 21:25:01	uksi-latest-arm64	docker	docker	>
matching-ws	10/03/2025, 21:24:35	uksi-latest-amd64	docker	docker	>
matching-ws	10/03/2025, 21:24:35	uksi-latest	docker	docker	>
matching-ws	11/03/2025, 05:35:20	taxref-latest-arm64	docker	docker	>
matching-ws	11/03/2025, 05:34:43	taxref-latest-amd64	docker	docker	>
matching-ws	11/03/2025, 05:34:44	taxref-latest	docker	docker	>
matching-ws	10/03/2025, 21:52:22	itis-latest-arm64	docker	docker	>
matching-ws	10/03/2025, 21:51:31	itis-latest-amd64	docker	docker	>
matching-ws	10/03/2025, 21:51:33	itis-latest	docker	docker	>
matching-ws	11/03/2025, 10:07:16	ipni-latest-arm64	docker	docker	>
sw-paniditem	2025-07-25T02:00:00Z	Panme-testable-ndqi	docker	docker	>
sw-paniditem	2025-07-25T02:00:00Z	tsajfe-siti	docker	docker	>
sw-paniditem	2025-07-25T02:00:00Z	Parame-testable-siti	docker	docker	>
sw-paniditem	2025-07-25T02:00:00Z	Parame-testable-siti	docker	docker	>
sw-paniditem	2025-07-25T02:00:00Z	Parame-testable-siti	docker	docker	>
sw-paniditem	2025-07-25T02:00:00Z	tsajfe-testable	docker	docker	>

Usage

```
docker run -p 8080:8080 docker.gbif.org/matching-ws:gbif-backbone-latest-amd64
```

```
docker run -p 8080:8080 docker.gbif.org/matching-ws:gbif-backbone-latest-arm64
```



Multi-taxonomy support in GBIF API



v2 Species match API

- `/v2/species/match?scientificName=Carcharodon+carcharias`
- `/v2/species/match?scientificName=Carcharodon+carcharias&checklistKey=2d59e5db-57ad-41ff-97d6-11f5fb264527`



WoRMS registry ID

v1 Occurrence API - search by name

- `/v1/occurrence/search?scientificName=Carcharodon+carcharia`
`s`
- `/v1/occurrence/search?scientificName=Carcharodon+carcharia`
`s&checklistKey=2d59e5db-57ad-41ff-97d6-11f5fb264527`



WoRMS registry ID

v1 Occurrence API - TaxonID

- `/v1/occurrence/search?taxonID=urn:lsid:marinespecies.org:taxname:225860&checklistKey=2d59e5db-57ad-41ff-97d6-11f5fb264527`



WoRMS registry ID

v1 Occurrence API - facets

- `/v1/occurrence/search?facet=familyKey`
- `/v1/occurrence/search?facet=subfamilyKey & checklistKey=7ddf754f-d193-4cc9-b351-99906754a03b`



CoL registry ID

v1 Occurrence download API

```
"predicate": {  
    "type": "equals",  
    "key": "TAXON_KEY",  
    "value": "urn:lsid:marinespecies.org:taxname:225860",  
    "checklistKey": "2d59e5db-57ad-41ff-97d6-11f5fb264527"  
}
```

SQL API

```
SELECT
    decimalLatitude, decimalLongitude, eventDate
FROM occurrence
WHERE
    array_contains(classifications ["2d59e5db-57ad-41ff-97d6-11f5fb264527"], "urn:lsid:marinespecies.org:taxname:158970")
    ↑
    registry ID
```

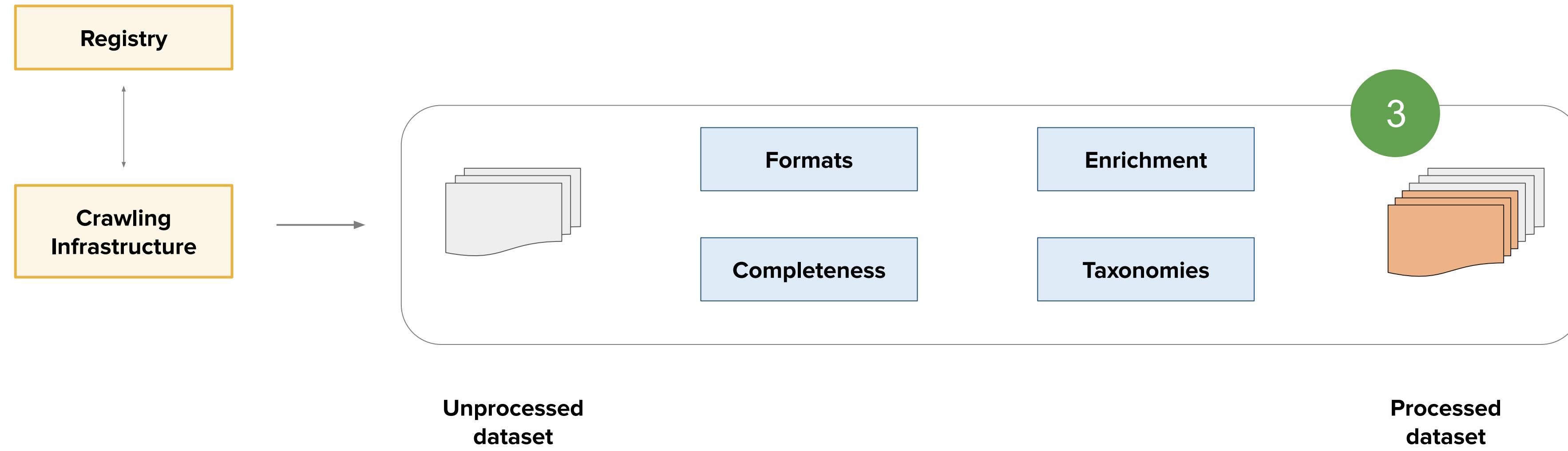
matching-ws

- “Catalogue of Life ready”
 - String identifiers
 - Support for intermediary ranks (eg. subphylum, subfamily)
- Consolidated ID support
 - WoRMS, IPNI, Dyntaxa, UK NBN
- Includes IUCN status
 - Scope for additional status values (e.g. invasive)

Processing as a service

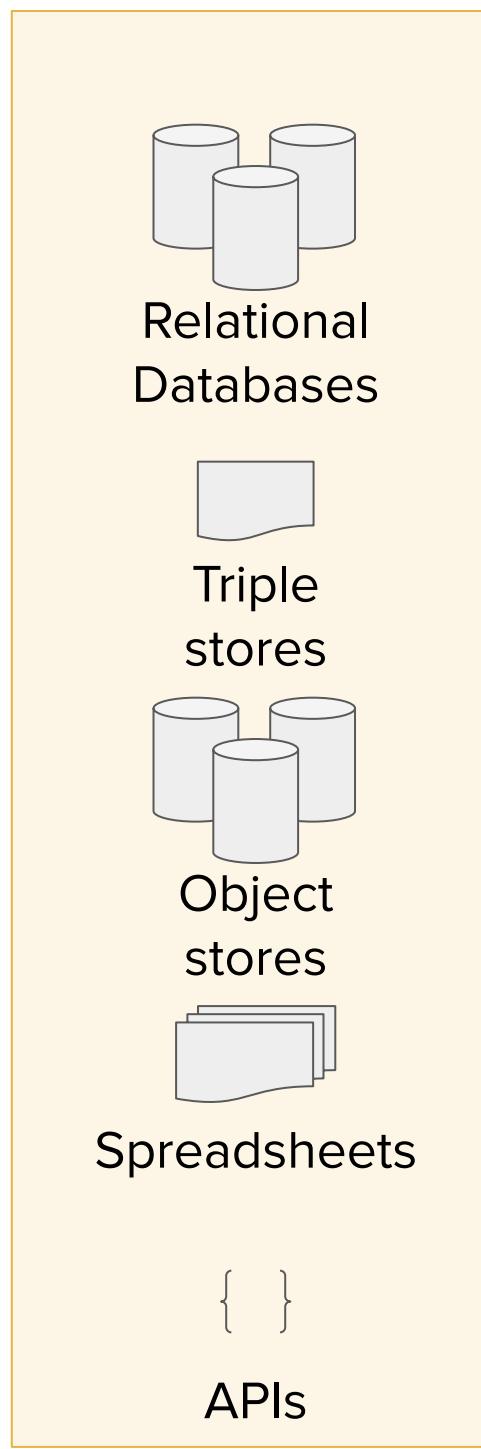


Idea: Processing as a service?

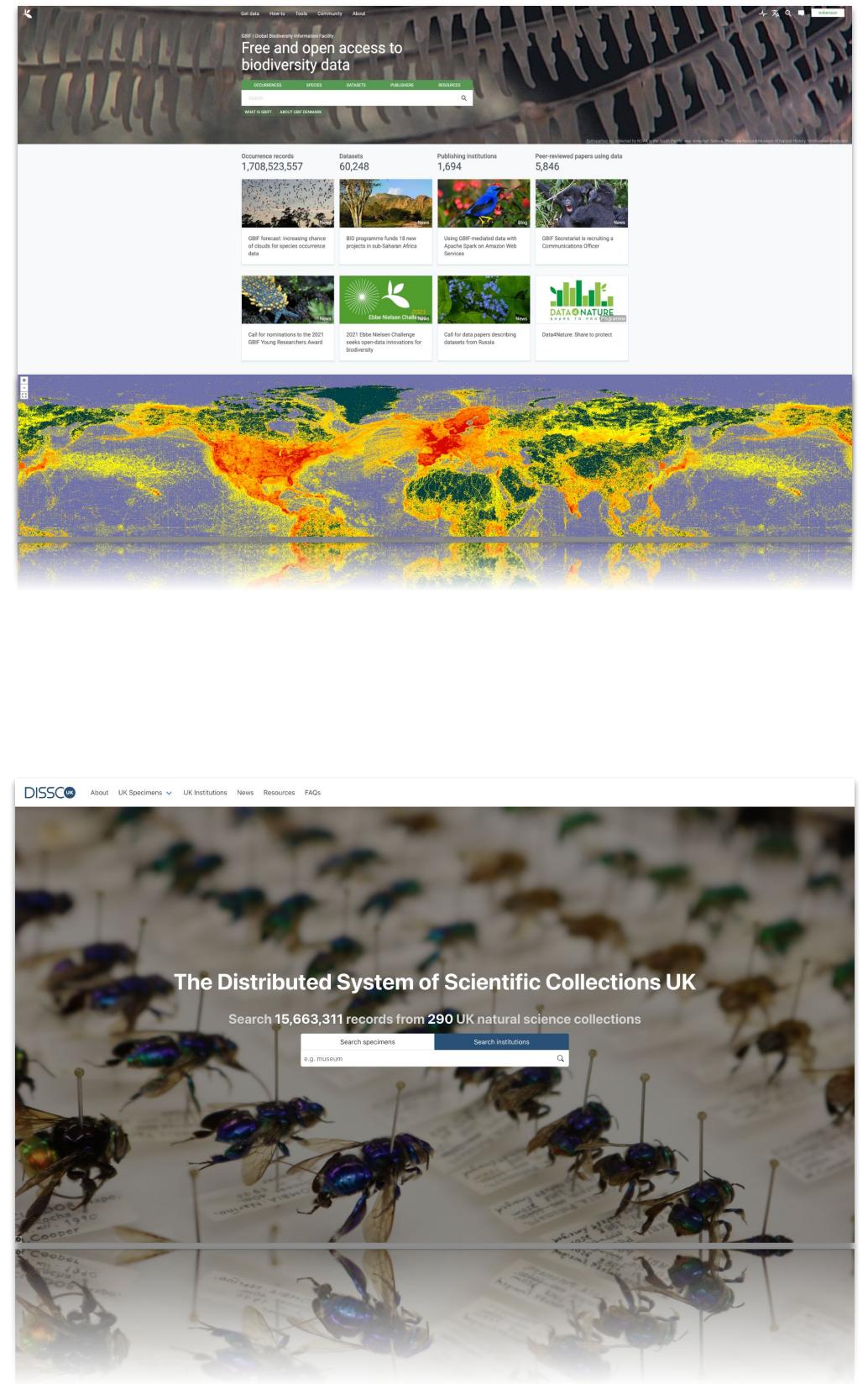
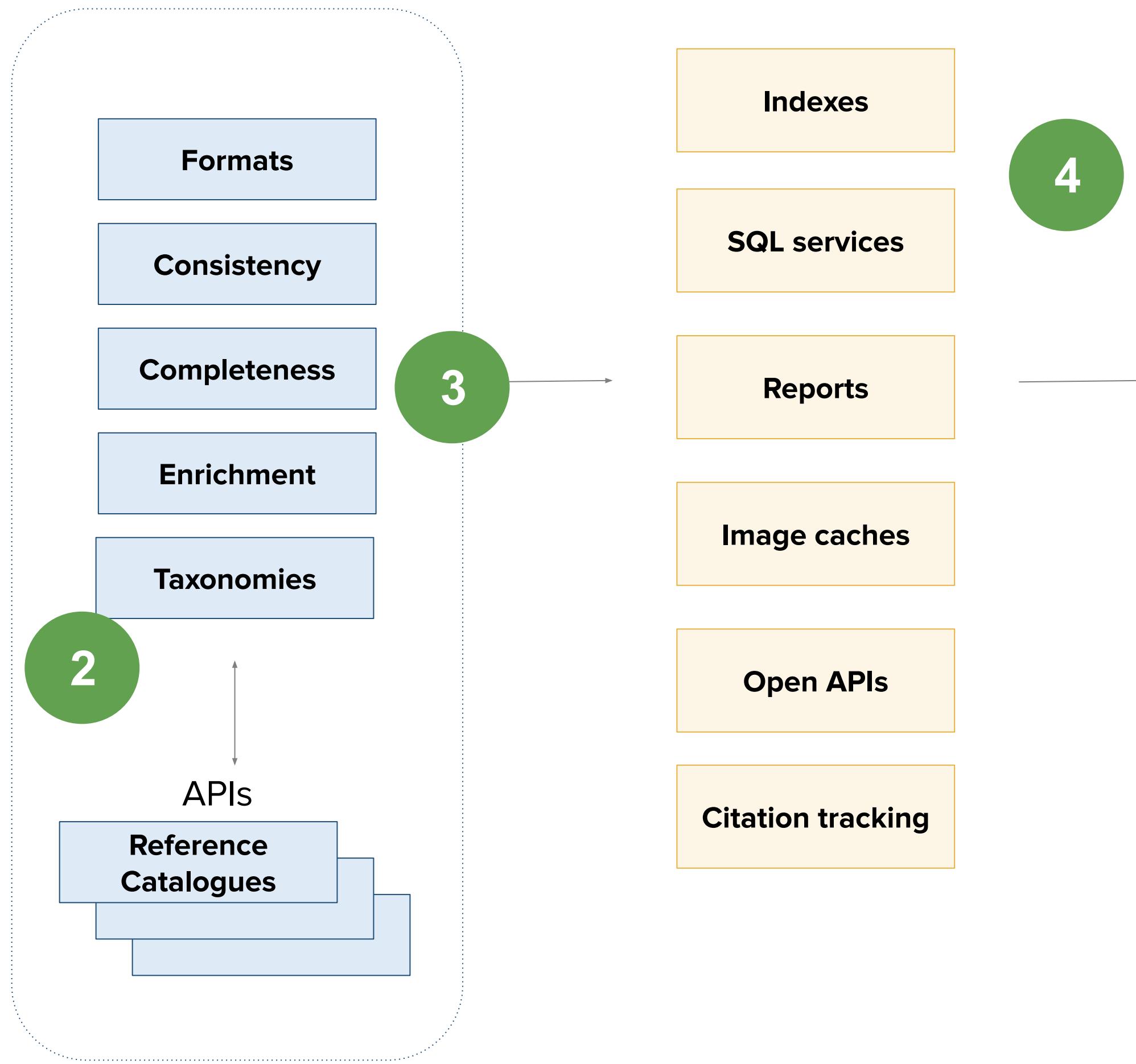


GBIF Data Warehouse (simplified)

Source Systems



Data pipelines



Thanks!

