

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 left is the ChecklistBank logo and a navigation menu with 'Datasets', 'Tools', and 'COL24.7'. A 'Login' button is in the top right. The main header features a banner with the ChecklistBank logo and the tagline 'Index and repository for taxonomic data' over a background image of yellow and orange flowers. Below the banner is a statistics table:

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 the table is a paragraph explaining the Catalogue of Life (COL) and ChecklistBank's role. To the right, a section titled 'Latest datasets added' lists three recent publications. At the bottom, there is a footer with development information and logos for GBIF and the Global Core Biodata Resource.

Open data, open access
taxonomic data publishing
platform

Tools for:

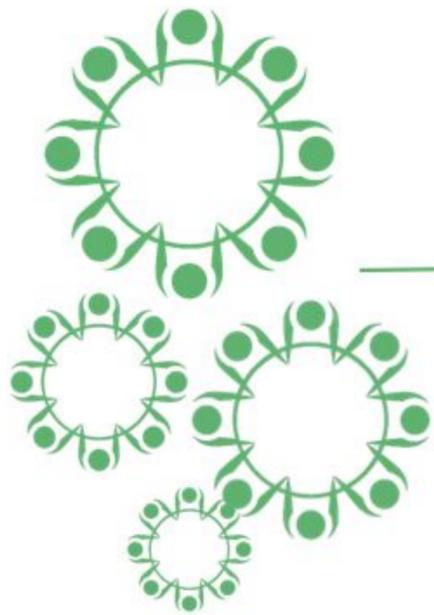
1. Importing lists
2. Comparing lists
3. Building lists

checklistbank.org

api.checklistbank.org



**Global
checklists**

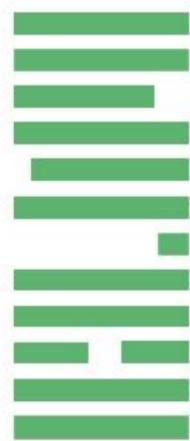


**Taxonomic
Communities**

**Regional /
management
checklists**



COL project



**Management classification*

**Attach sectors*

**Editorial decisions*

**Persistent name ID's*



Release

base

→ **Established**

→ **Brand new**

→ **Coming soon**

Catalogue of Life



*COL checklist as
a source*



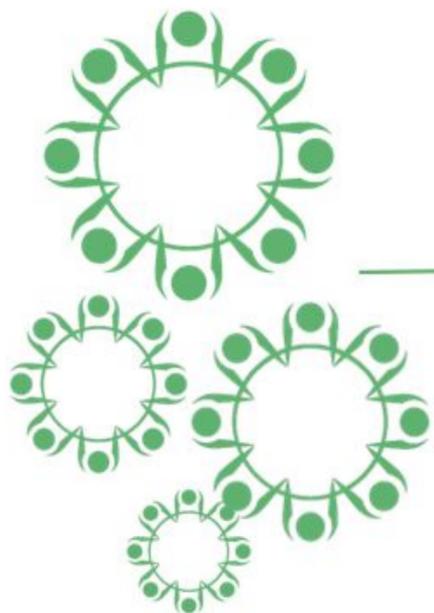
GBIF

**Backbone
Taxonomy**

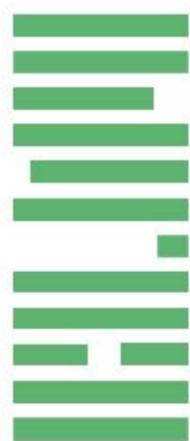


ChecklistBank

**Global
checklists**



COL project



**Management classification*

**Attach sectors*

**Editorial decisions*

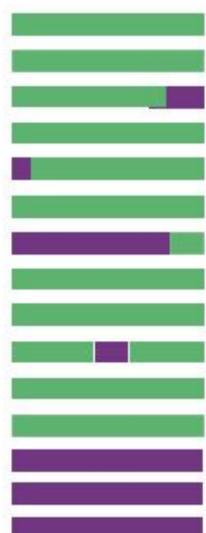
**Persistent name ID's*



Release

COL project

base



** Merge sectors*

** family and below*

** Homotypic grouping*

** Editorial decisions*

** Quality content check*



Extended Release

→ **Established**

→ **Brand new**

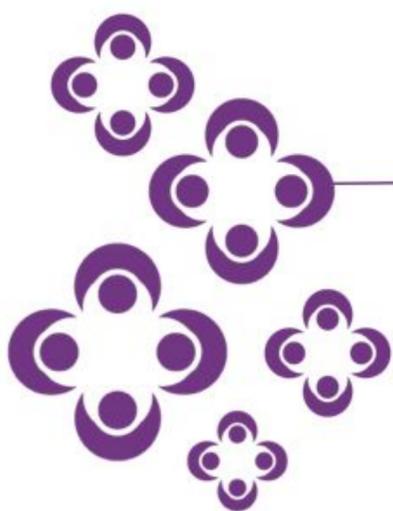
→ **Coming soon**

Catalogue of Life



COL checklist as
the **ONLY** source

**Regional/
management
checklists**

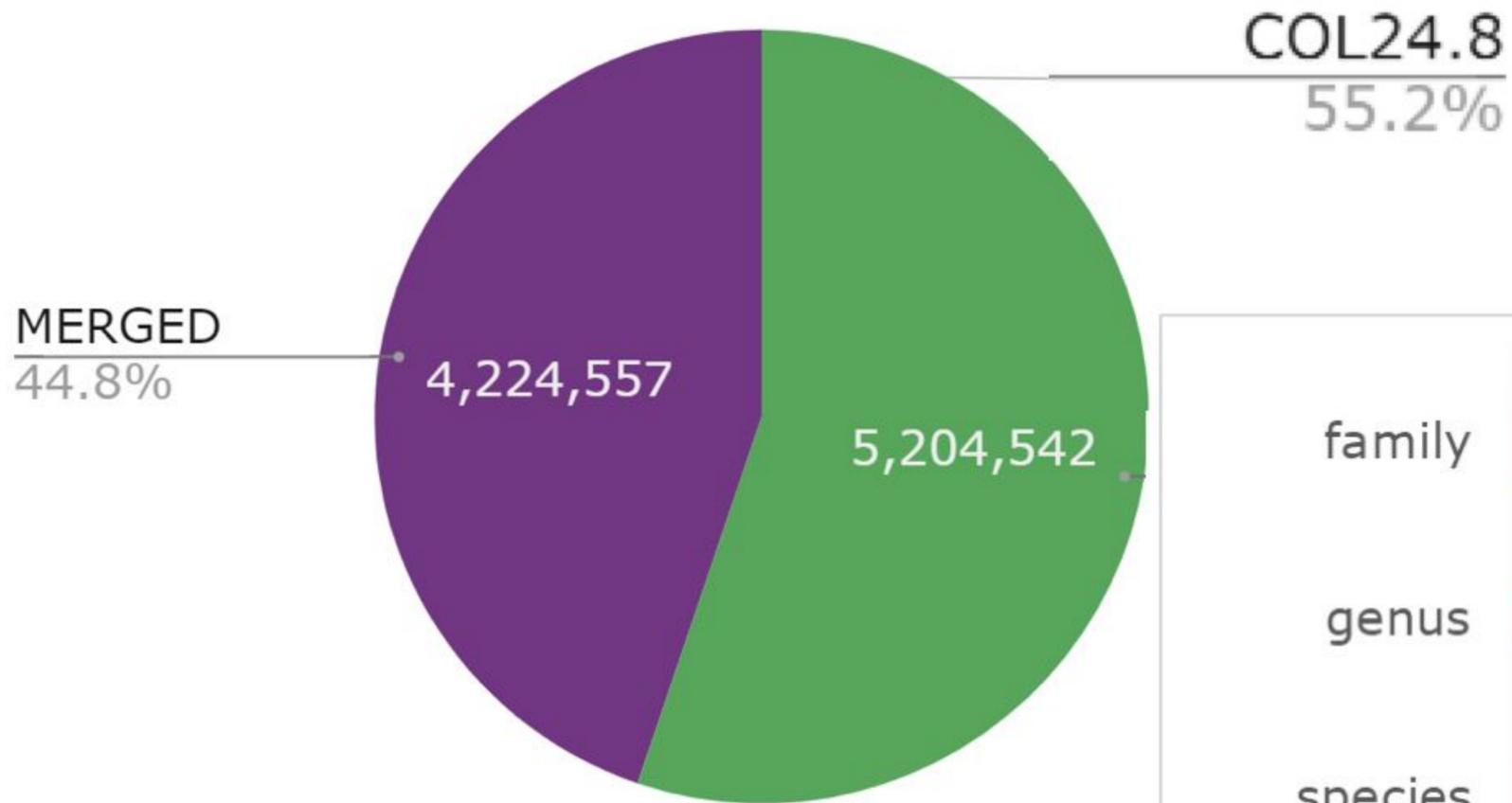


**Taxonomic
Communities**

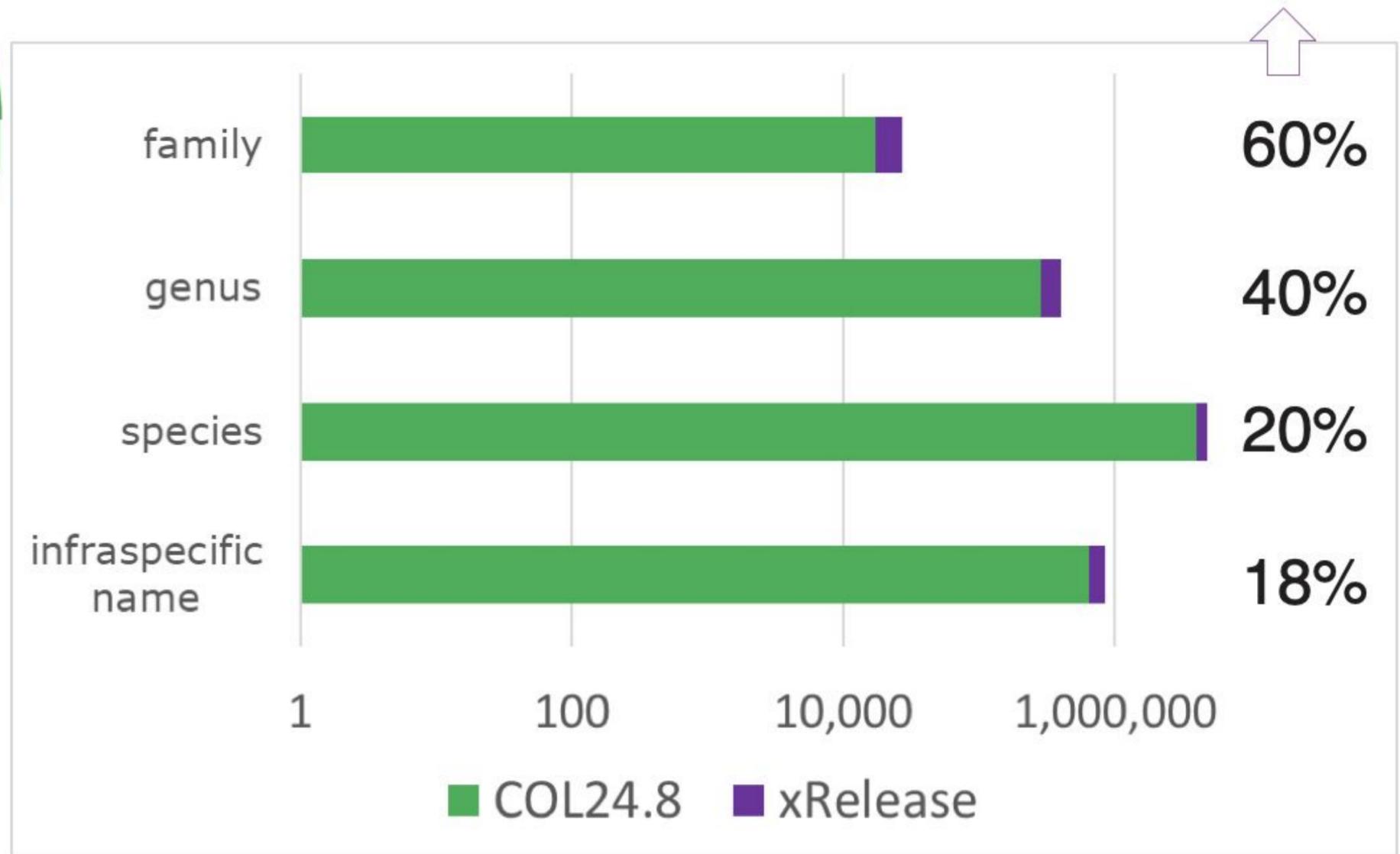
Merged names in COL (all ranks)

xRelease
2024-08-25

Name count

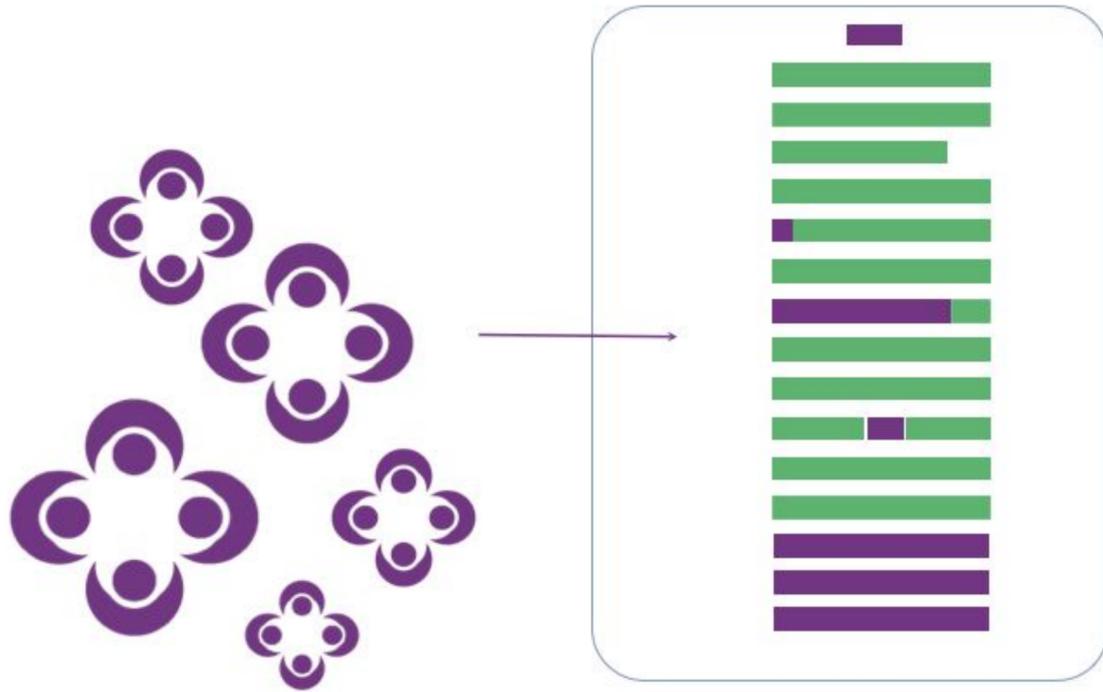


ONLY family names and below are being merged



You can contribute to a better COL

1. Publish or recommend a checklist



2. Issues report

Missing Chrysididae species #622

Open mdoering opened this issue last month · 3 comments

mdoering commented last month

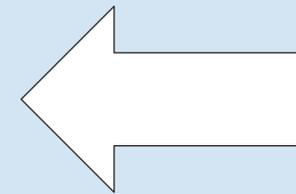
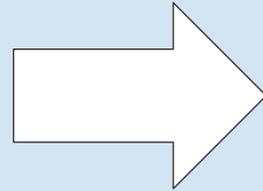
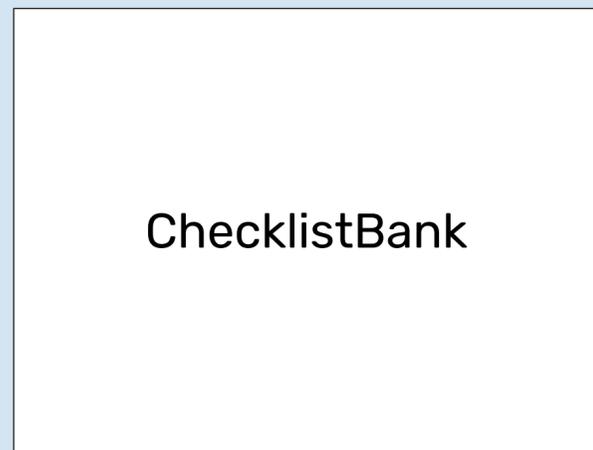
There are quite a few species missing from the [Chrysididae](#) family in COL 24.1

The following species were existing in the SANBI species dictionary, but not found in COL:

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

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

Matching-ws



Matching-ws



Matching-ws
(Catalogue of Life)

Matching-ws application

Main
Index
(names +
classifications)

Status info

IUCN

GISIN ?

identifiers

WORMS

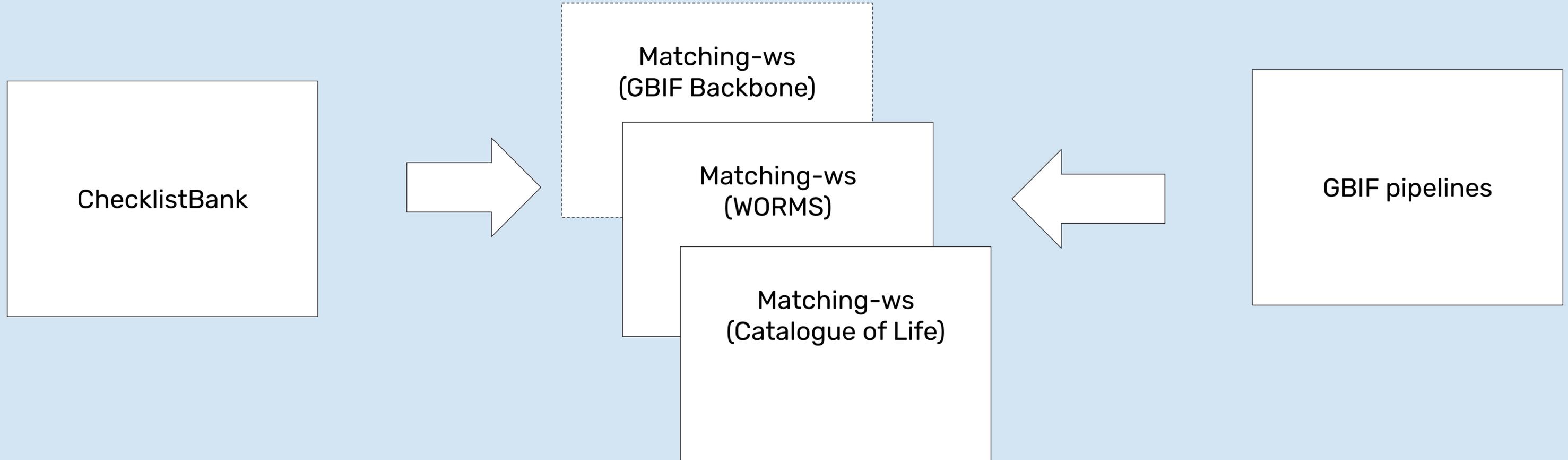
IPNI

dyntaxa

UK SI



Matching-ws



https://repository.gbif.org

The screenshot shows the Sonatype Nexus Repository web interface. The browser address bar displays 'repository.gbif.org/#'. The page header includes the Sonatype logo, 'Sonatype Nexus Repository OSS 3.75.1-01', a search bar with 'Search components', and navigation icons for refresh, help, and sign in.

The left sidebar is titled 'Browse' and contains a list of categories: Welcome, Search, Custom, Docker (highlighted in green), Helm, Maven, NuGet, RubyGems, npm, and Browse.

The main content area is titled 'Q Docker Search for components in Docker repositories'. It features search filters for 'Image Name' (matching-ws), 'Image Tag' (Any), 'Layer Id' (Any), and 'Content Digest' (Any). A '+ More criteria' dropdown is also visible.

Below the filters is a table of search results:

	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	>



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+carcharias`
- `/v1/occurrence/search?scientificName=Carcharodon+carcharias&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-97d6-11f5fb  
264527"], "urn:lsid:marinespecies.org:taxname:158970")
```

registry ID



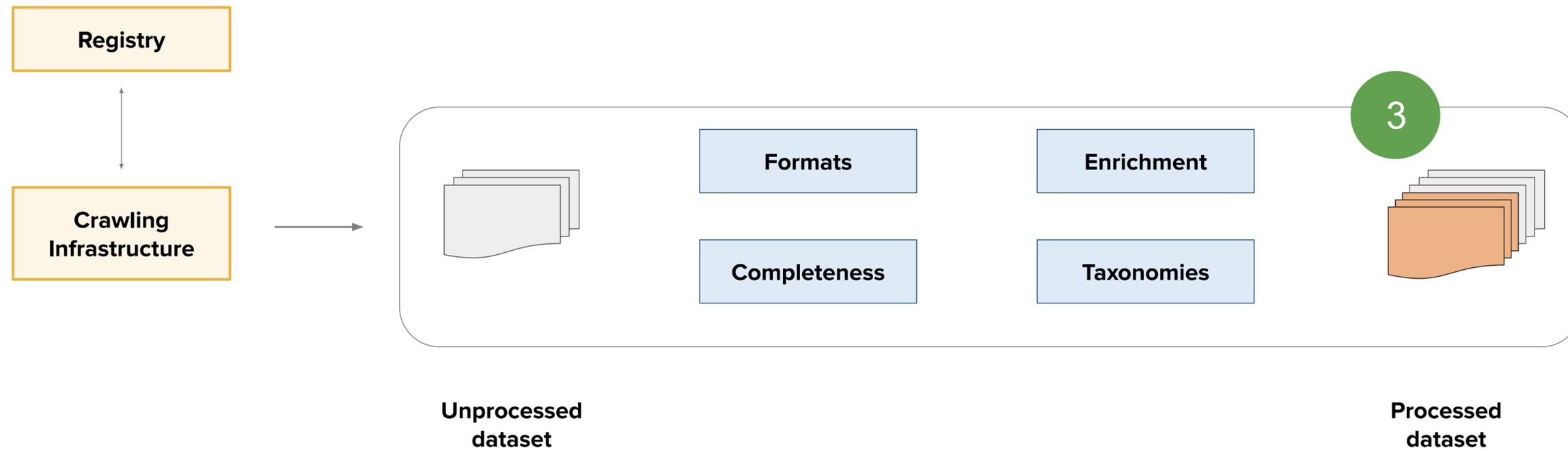
matching-ws

- “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

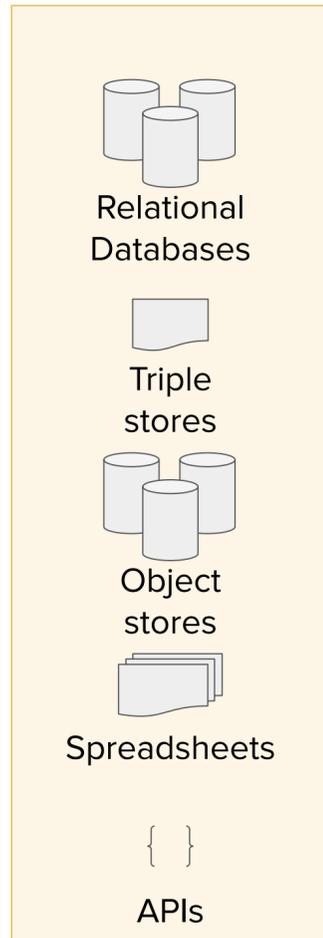


Idea: Processing as a service?

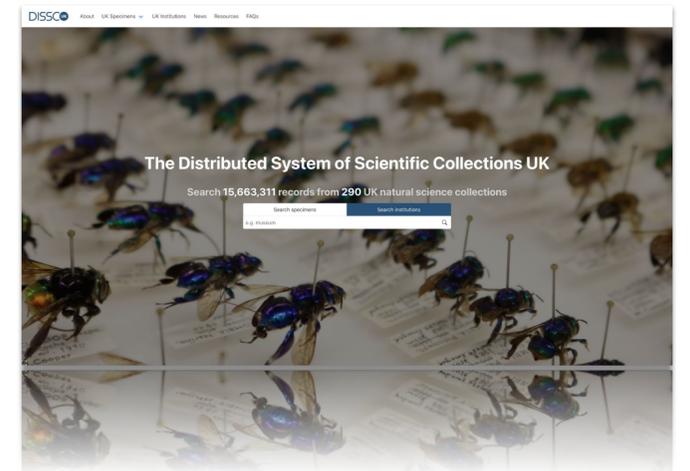
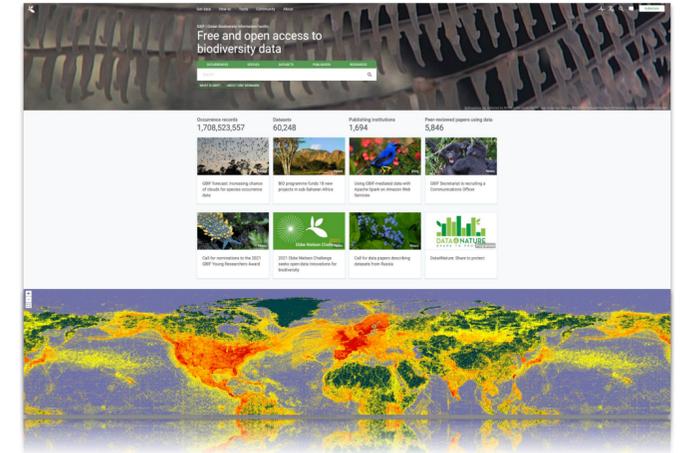
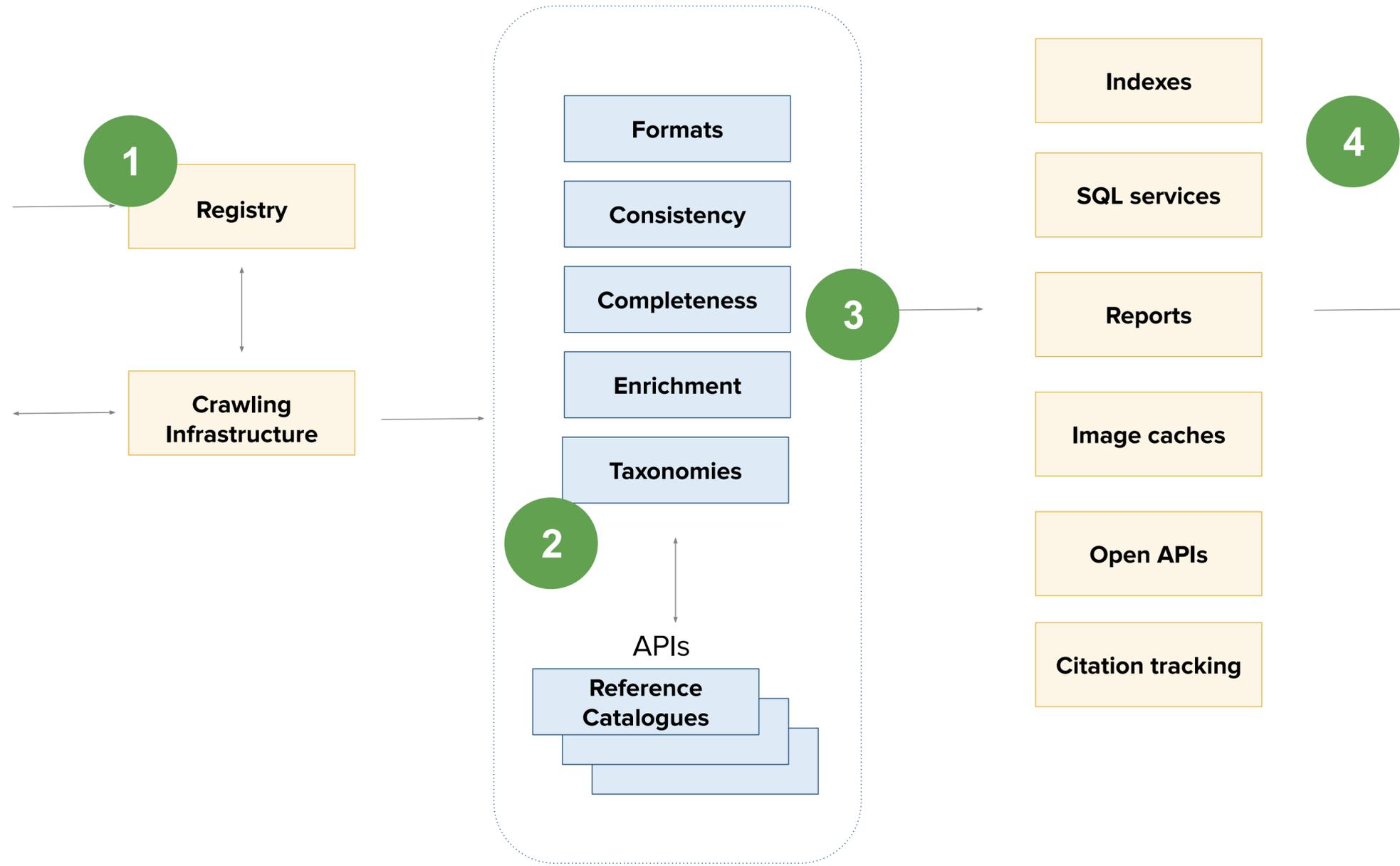


GBIF Data Warehouse (simplified)

Source Systems



Data pipelines



Thanks!

