

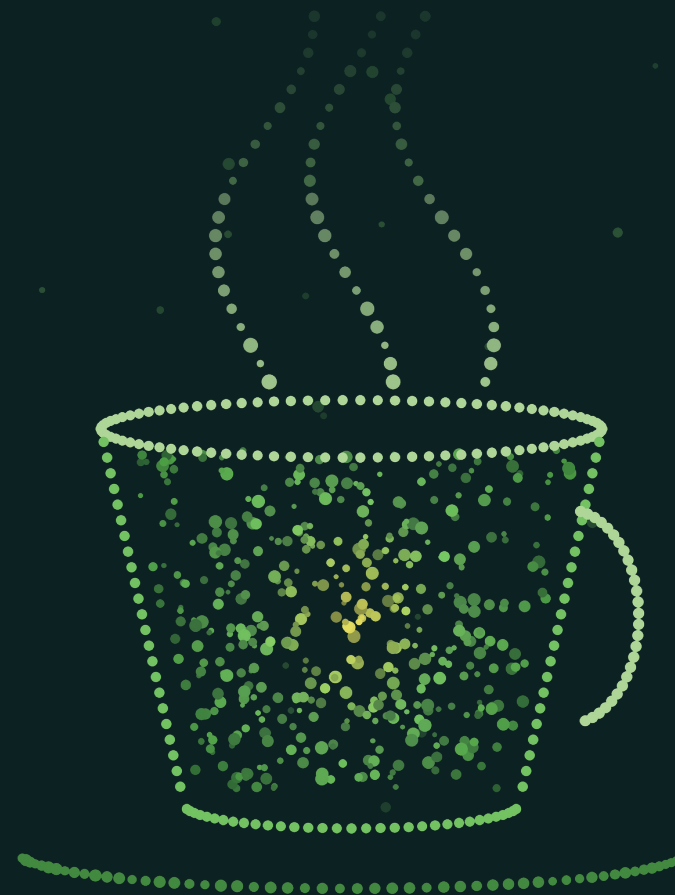
GBIF SWEDEN

SESSION 2 · 9 OCTOBER 2026

REUSING DATA FROM GBIF

THE WEBSITE, R, CUBES AND THE DOI

Lena Thöle

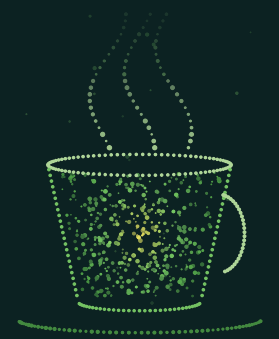


Before we start



- **We are recording.** The recording and the slides go on gbif.se/data-fika afterwards.
- **Microphones & cameras are off.** Type questions in the chat and just relax.
- **Questions in the chat at any time.** We take them as they come, and again at the end.
- **Poll.**

The ten sessions – one journey



The harlequin ladybird



Is the harlequin ladybird spreading in Sweden?

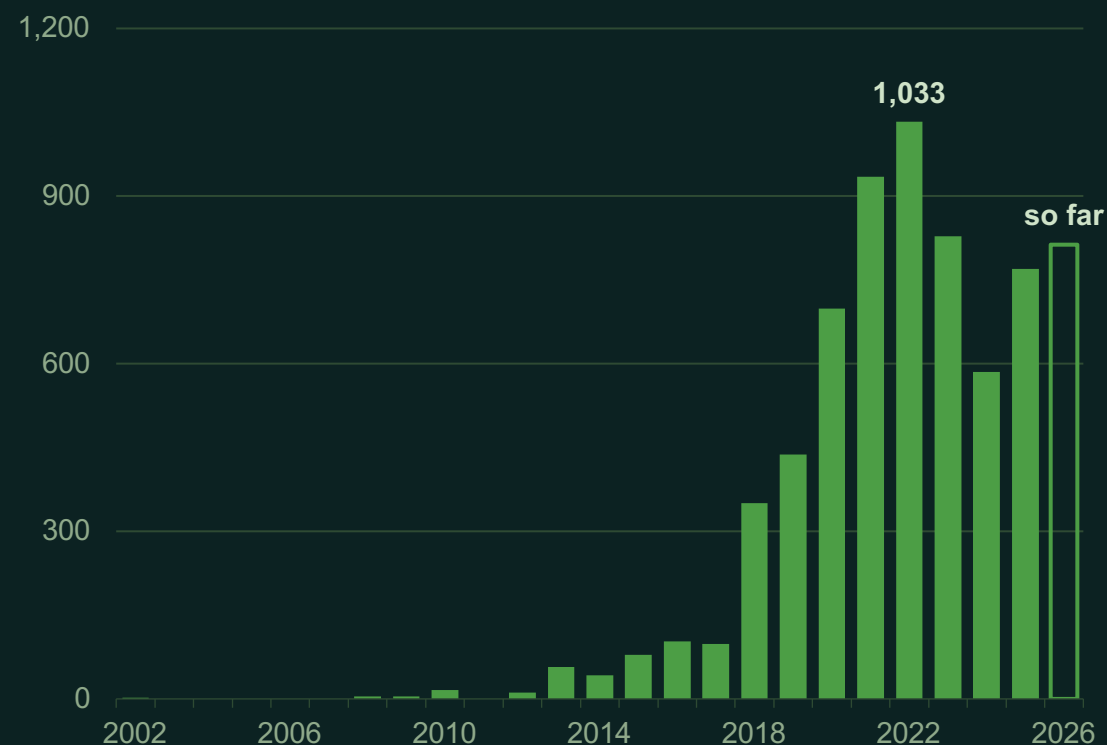


Harmonia axyridis · harlekinpiga

6,867

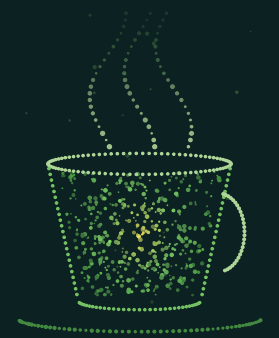
records from Sweden in GBIF, 5 October 2026

- 2 records in 2002, more than 1,000 in 2022
- Today: get them three ways, and end every way with a citation



Records per year in Sweden · GBIF.org, Catalogue of Life · 2026 so far

Today



0

Prerequisites

GBIF account

1

The website

Which door, your account, the filters that matter, what is in the zip

2

Code

The same download in R, and how it looks in Python

3

Cubes

Counts instead of records

4

Cite it

Where the DOI goes, and derived datasets

One example all the way through: the harlequin ladybird in Sweden

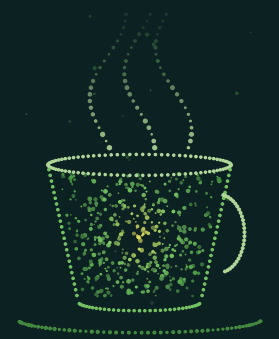
0

The prerequisites



Where to find what?

Updated weekly



GLOBAL AND SWEDISH

GBIF.org

- A DOI for every download
- Rgbif & pygbif
- Cubes
- Catalogue of Life backbone
- Collections from all over the world

Default for anything you publish

SWEDISH

SBDI

- Python: galah; (R: sbdi4r2)
- Swedish tools, e.g. the ASV portal for DNA data
- New things coming soon:
 - Portal in Swedish
 - Dyntaxa use



Swedish questions

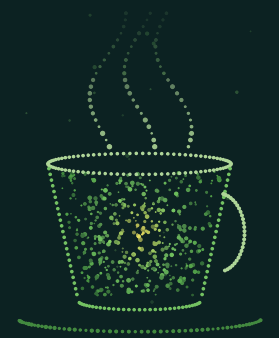
A MAIN (CITIZEN SCIENCE) DATA SOURCE

Artportalen

- Main portal for Swedish Citizen Science observations
- Sensitive species on application (data fika session #7)

The latest records

Your GBIF account



- Log in required for download
- ORCID, GitHub or Google
- The same account runs rgbif, pygbif, cubes and derived datasets
- Your Downloads page keeps the query, the DOI and the deletion date

Every download credits the datasets inside it.

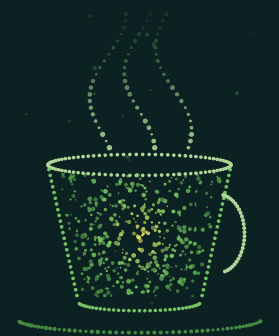
That is how Artportalen, iNaturalist, research data or museum collections see their data being used.

01

The website



Filters that change your answer



What, where, when

Scientific name

Country or area

Year

matched to the Catalogue of Life

What kind

Basis of record

Occurrence status

absences are records too

How good

Has coordinates

Geospatial issues

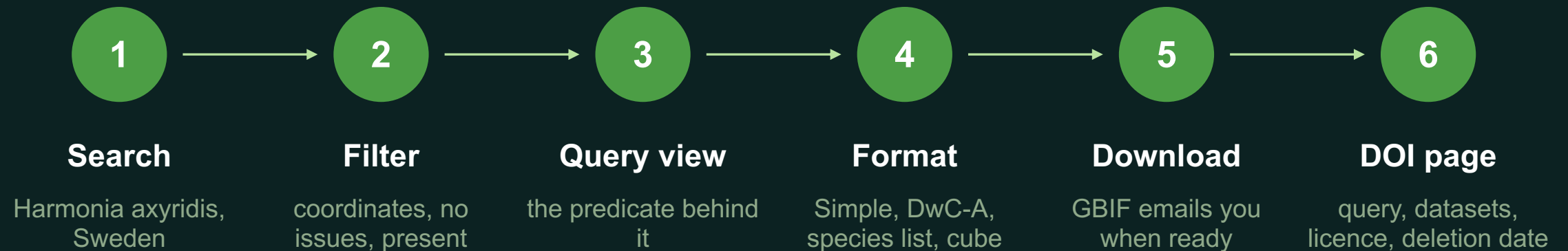
Coordinate uncertainty

Reuse

Licence

the download takes the strictest licence inside it

From search to DOI



Seen in session #1: search, filters and the dashboard. New today: the query view and the formats.

What's in the zip?



Four formats

Simple CSV

one row per record, about 50 columns.

Darwin Core Archive

everything: interpreted, verbatim, multimedia, metadata

Species list

just the names

Cube

counts, not records (part 3)

Occurrences, events and DwC-DP

- A download is a flat table of occurrences, mostly presence-only.
- Event data also says where people looked and how hard: protocol, effort, sometimes absences.
- DwC-A is today's zip, DwC-DP, ratified in May 2026, keeps events, occurrences and measurements linked.

What the columns mean: session #3

The data model: session #4

02

Code

The same download as a script:
re-runnable, citable, and the same in R and Python.



Search to explore, download to cite



FOR LOOKING

`occ_search()`

- Instant
- Max 100,000 records
- No DOI
- Results change as the data change

FOR YOUR PAPER

`occ_download()`

- Any size
- A DOI, with the query stored
- The same file for everyone, later
- Takes a few minutes

Package description:

<https://www.gbif.org/tool/81747/rgbif>

Set up once

R · .Renviron

```
install.packages(c("rgbif", "usethis"))
usethis::edit_r_environ()

# add three lines, save, restart R
GBIF_USER="your_username"
GBIF_PWD="your_password"
GBIF_EMAIL="you@institution.se"

# never in a script, never in git, never on a screen share
# Python (pygbif) reads the same three variables
```

The whole download

R · rgbif

```
library(rgbif)

tx <- name_backbone("Harmonia axyridis")
tx$usageKey                                # "PNVYQ"

dl <- occ_download(
  pred("taxonKey", tx$usageKey),
  pred("country", "SE"),
  pred_default(),                          # coordinates, no geo-issues, present
  format = "SIMPLE_CSV")

occ_download_wait(dl)
d <- occ_download_get(dl) |> occ_download_import()
gbif_citation(occ_download_meta(dl))$download
```

Same thing in Python

PYTHON · pygbif

```
from pygbif import species, occurrences as occ

tx = species.name_backbone(scientificName="Harmonia axyridis")
key, _ = occ.download([f"taxonKey = {tx['usage']['key']}",
                        "country = SE",
                        "hasCoordinate = TRUE"],
                        format="SIMPLE_CSV")

occ.download_meta(key) ["status"]      # wait for SUCCEEDED
occ.download_get(key, path="data")
print(occ.download_citation(key))
```

One API behind every door

YOU USE

GBIF.org
the website

rgbif
R

pygbif
Python

Any web request
curl, a browser, any language

IT CALLS

The GBIF API

`api.gbif.org/v1`

One address, one set of
filters, one copy of the data

WHICH ANSWERS WITH

Search `/occurrence/search`
Explore · instant · max 100,000 rows
`occ_search()`

NO LOGIN

Download `/occurrence/download`
Cite · zip file · gets a DOI
`occ_download()`

LOGIN

Cube `/occurrence/download/request/sql`
Counts per species, place, year · DOI
`occ_download_sql()`

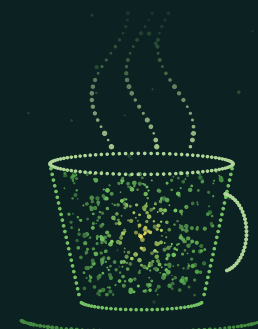
LOGIN

Names `/species`
Turns a name into a taxon key
`name_backbone()`

NO LOGIN

Same data, same filters, same DOI — whichever door you use. The LOGIN rows are why R and Python need your GBIF account.

GBIF changed its taxonomy



4989904 → PNVYQ

Harmonia axyridis: old GBIF backbone key →
Catalogue of Life key

6,933 vs 6,867

Swedish records, same filters: old backbone vs
Catalogue of Life

- Since June 2026 GBIF.org uses the Catalogue of Life, and the old backbone is frozen.
- Latest packages have now default COL.

03

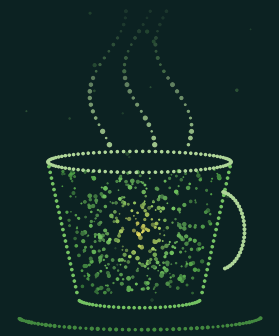
Cubes

Counts instead of records: small enough to email, and it still has a DOI.

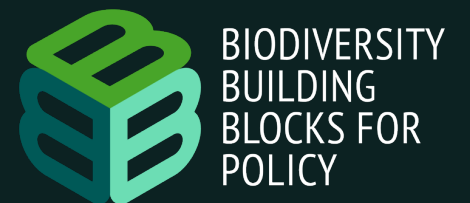
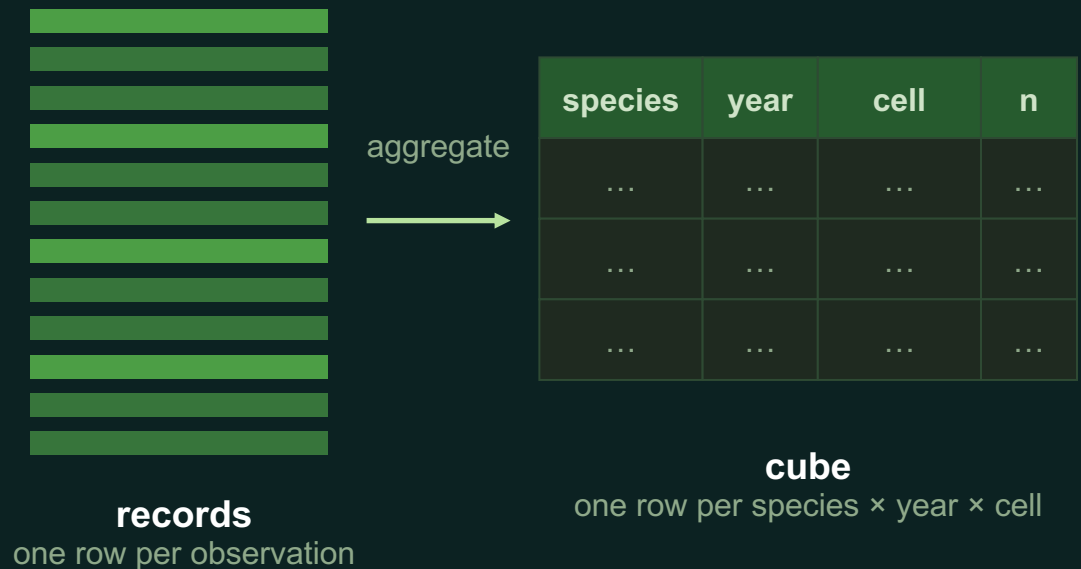


What's a cube?

- **Counts instead of records:** species × time × grid cell → number of occurrences
- Plus the smallest coordinate uncertainty in each cell
- **Tiny:** 1.5 million European lagomorph records became a 23 KB cube
- Gets a DOI like any download.
- Developed with the **EU B-Cubed** project:
 - Website: <https://b-cubed.eu/>
 - Tutorials: <https://docs.b-cubed.eu/guides/occurrence-cube/>



Structure of a GBIF species occurrence cube



The ladybird cube



1

GBIF.org: Download → Occurrence cube → Configure. Species × year × EEA 10 km grid.

2

Click “Edit as SQL” and copy the query.

3

In R: `occ_download_sql(sql)` gives a zip and a DOI, like any download.

```
SELECT "year", species,  
       GBIF_EEARGCode(10000, decimalLatitude,  
                       decimalLongitude,  
                       COALESCE(coordinateUncertaintyInMeters,  
                                1000)) AS eeaCellCode,  
       COUNT(*) AS occurrences  
FROM occurrence  
WHERE countryCode = 'SE'  
       AND species = 'Harmonia axyridis'  
       AND occurrenceStatus = 'PRESENT'  
GROUP BY "year", eeaCellCode, species
```

From “Edit as SQL” on GBIF.org · run from R with `occ_download_sql()`

04

Cite it

#citethedoi



Cite it right



One download, one DOI

```
GBIF.org (9 October 2026) GBIF Occurrence  
Download https://doi.org/10.15468/dl.xxxxxx
```

- In the reference list, not only in the methods
- It credits every dataset inside: Artportalen, iNaturalist, Museum, research data, ...
- Uncited, the zip can be deleted after six months. The DOI and the query stay.

Filtered, cleaned or combined?

Register a derived dataset: the datasets and record counts you actually used get their own DOI.

```
derived_dataset(  
  citation_data = count(d_clean, datasetKey),  
  title = "Ladybirds in Sweden, cleaned",  
  description = "Cleaned: see Methods",  
  source_url = "https://zenodo.org/records/...")
```

#citethedoi

Only a download has a DOI, and only a DOI lets someone check your result.



YOUR TURN

One download of your own

Log in to GBIF.org (R and Python need a GBIF password).

Download your own taxon or place, in the format you need.

Last workshop places are open now: gbif.se/data-fika



Questions?

Lena Thöle · lena.thole@nrm.se · gbif@nrm.se

Slides, recording and the demo script: gbif.se/data-fika



Next fika

Session 3 · Friday 13 November, 11:00–11:45

Standards, Identifiers & Metadata

What the columns in today's download actually mean.

Thank you for the fika – see you next month.



gbif.se/data-fika

Slides, recording and all materials

Links

- [GBIF Website](#)
- [GBIF Sweden](#)
- [SBDI](#)
- [Artportalen](#)
- [GBIF data use club](#)
- [Rgbif package description](#)
- [Pygbif package](#)
- [B Cubed website](#)
- [B-Cubed Tutorials](#)

