Q&A log from 20 February 2026

Creating species occurrence cubes with GBIF

How do I sign up to receive a certificate? I didn't know that was an option

- Please send an email to info@b-cubed.eu with your names and the sessions you attended and we will send you the certificate.

Is it possible to have these slides?

- We will send the slides a couple of days after the session, along with a recording of the training.

In this case you will get also absence data

- Yes, you can get absence data if they are available. You can use the "occurrenceStatus column" for this purpose. Some options are i.e. filter for PRESENT, for ABSENT, or include it in the SELECT and GROUP BY to get both.

Are these species keys from the gbif backbone, or from the catalogue of life (or are both possible through the sql api)?

- They are from the GBIF backbone

Why is GBIFgiving preference to the species key instead of the botanical name? Does it include synonyms or recently changed/shifted names.

- 'speciesKey' includes synonyms, subspecies, and synonyms of subspecies. 'genus' and 'specificEpithet' would be needed to get the name. The full list of columns is at https://techdocs.gbif.org/en/data-use/api-sql-downloads#sql-columns

What exactly is a non-null value? Meaning no absence data?

- A null value is the absence of a value, e.g. a record in the ocean with no country code. Not-null / non-null means the value can't be null, for example the columns gbifid and datasetkey are never null. In other words, it counts fields that do have (any) value

In aggregating functions, is it possible to set more than one country code/years etc. in a single query?

- Yes, this is possible. You can specify different country codes, years, time span, etc.

I just wonder, whether there is any chance these very simple counting queries could be used without "waiting" for the download?

- No. The original occurrence API can do immediate counting for many queries, e.g. count by dataset for records in Guernsey: https://api.gbif.org/v1/occurrence/search?facet=datasetKey&facetLimit=150000&limit=0&country=GG. One day some sort of preview for SQL queries might be possible, in a similar way to Google BigQuery.

Why are there closed parentheses after 2026?

- Good catch. It should not be there.

Does the order of GROUPBY column names affect the query result?

- No, it does not

Can anyone link me to research articles that use this method or this tool, if there are any?

- We are working on them at the moment. There are several deliverables already available at https://www.b-cubed.eu/library.

Can the queries be run to extract values with a range of gridsquares?

- Yes, you can do it for certain grids.

I cannot find information regarding the field 'invasive' - is the status assigned on a country level and can the source of the information be tracked somehow?

- The field is as follows: occurrence.establishmentmeans.concept IN ('INVASIVE', 'NATURALISED')

In the GBIF Darwin Core Type Vocabulary, the options for basisofrecord are given as eg MachineObservation not 'MACHINE_OBSERVATION'. It's also not mentioned in the table with available columns in the SQL download API docs. How can we know the exact terms to be used?

- This is a good point, and not very visible. They're shown in the enumeration API: http://api.gbif.org/v1/enumeration/basic/BasisOfRecord. Or on an individual record: https://api.gbif.org/v1/occurrence/1424468740

Hi everyone!! Does it make any difference in the output, the order in which we write the filters in WHERE?

- No really

The cube looks like what they use for the EBVs calculation

- Data cubes are a good approach for assessing multidimensional data sets as EBVs. Particularly, B-Cubed is also collaborating with the calculation of EBVs.

Is data cube somewhat similar to data matrix?

- Indeed. The result is stored in a tabular format, and contains the specified fields of the data aggregation with the SQL specifications

Does the result of the spatially chosen point within the uncertainty change with consecutive queries? If yes, I will potentially obtain different occurrences on each grid every time, right?

- Yes, that's right.

Ok, this can cause huge problems for a company that would like to understand changes in species occurrences after an intervention. How can this uncertainty be decreased?

- All data, including GPS points, have uncertainty. You can set the uncertainty range that you prefer for selecting the observations. In terms of the uncertainty for the cubes generations, you can turn it off.

Are there standard sets of filters? Can we create our own and share these?

- You can use the filters based on JAVA language and the DWC vocabulary