BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 4: e59154
doi: 10.3897/biss.4.59154

OPEN ACCESS

Conference Abstract

# Occurrence Cubes: A new way of aggregating heterogeneous species occurrence data

Damiano Oldoni‡, Quentin Groom§, Peter Desmet‡

‡ Research Institute for Nature and Forest (INBO), Brussels, Belgium
§ Meise Botanic Garden, Meise, Belgium

Corresponding author: Damiano Oldoni (damiano.oldoni@inbo.be)

Received: 30 Sep 2020 | Published: 30 Sep 2020

Citation: Oldoni D, Groom Q, Desmet P (2020) Occurrence Cubes: A new way of aggregating heterogeneous species occurrence data. Biodiversity Information Science and Standards 4: e59154.
https://doi.org/10.3897/biss.4.59154

## Abstract

The digital era has brought about an impressive increase in the volume of published species occurrence data. Research infrastructures such as the Global Biodiversity Information Facility (GBIF), the digitization of legacy data, and the use of mobile applications have all played a role in this transition. More data implies, unavoidably, more heterogeneity at multiple levels as a result of the different methods and standards used to collect data. Data standardization and aggregation help to reduce this heterogeneity. Furthermore, intermediate data products that can be used for activities such as mapping, modeling and monitoring improve the repeatability and reproducibility of biodiversity research (Kissling et al. 2017).

Occurrences can be defined as events in a three-dimensional space where the dimensions are taxonomic (what), temporal (when) and spatial (where). They are then aggregated into what we coined *occurrence cube* (Fig. 1).

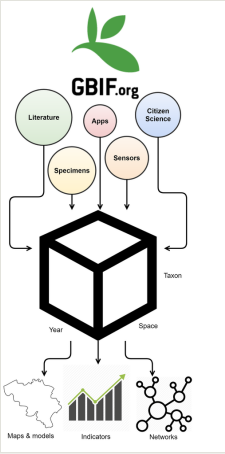


Figure 1.

Occurrence cubes (centre) are the result of aggregating species occurrence data from different sources and at different scales by taxon, time and space. It can be used as a data product for maps, models, indicators or networks.

The taxonomic dimension is categorical. Research infrastructures like GBIF use a taxonomic backbone, thus making data aggregation at species level or higher rank relatively easy. The temporal dimension is a continuum and the temporal uncertainty is usually lower than the typical aggregation span, typically a year. Regarding the spatial dimension, occurrences are typically filtered to remove those with too large an uncertainty to fit the grid scheme being used. Meaning that the spatial uncertainty is largely unused. We developed a method to take into account this spatial uncertainty while aggregating data. In particular, we state that an occurrence is spatially representable as a closed plane figure such as a circle, hexagon or square, never as the geometric centre (centroid) of it. As for GBIF occurrence data, the *coordinateUncertaintyInMeters* is defined as the radius describing the smallest circle containing the whole of the location (see Darwin Core standard). So, spatially speaking, we refer to occurrences as circles, even if the method described below is general.

After harvesting the occurrence data and providing a data quality assessment (e.g. removing occurrences without coordinates or with suspicious coordinates) we can assign occurrences to a reference grid such as the European reference grid of the European Environment Agency (EEA) at 1 km scale. In this spatial aggregation we randomly choose a point within the occurrence circle and assign it to the grid cell in which it is contained. We can aggregate further by time (e.g. by year) and taxonomy (e.g. by species), where aggregating means *counting* how many occurrences are in each specific taxonomic-spatial-temporal unit.

The analogy with geometry goes further: the occurrence cube can, as any cube, be projected on an orthogonal plane by aggregating along one of the three dimensions. In particular, projecting the cube on the taxonomic and temporal dimensions can be done by

adding up the number of occurrences, or counting the number of occupied cells, thus estimating the area of occupancy.

The occurrence cube paradigm has been developed within the Tracking Invasive Alien Species (TrIAS) project (Vanderhoeven et al. 2017) following Open Science and FAIR principles. We created and published occurrence cubes at the species level for Belgium and Italy (Oldoni et al. 2020b) and the occurrence cubes for non-native taxa in Belgium and Europe (Oldoni et al. 2020a).

## Keywords

methodology, data analysis, GBIF, species distribution modelling, indicators, TrIAS

## Presenting author

Damiano Oldoni

## Presented at

TDWG 2020

## Funding program

BelSPO BR/165/A1/TrIAS

## References


- Kissling WD, Ahumada J, Bowser A, Fernandez M, Fernández N, García EA, Guralnick R, Isaac NB, Kelling S, Los W, McRae L, Mihoub J, Obst M, Santamaria M, Skidmore A, Williams K, Agosti D, Amariles D, Arvanitidis C, Bastin L, De Leo F, Egloff W, Elith J, Hobern D, Martin D, Pereira H, Pesole G, Peterseil J, Saarenmaa H, Schigel D, Schmeller D, Segata N, Turak E, Uhlir P, Wee B, Hardisty A (2017) Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. Biological Reviews 93 (1): 600-625. https://doi.org/10.1111/brv.12359
- Oldoni D, Groom Q, Adriaens T, Davis AS, Reyserhove L, Strubbe D, Vanderhoeven S, Desmet P (2020a) Occurrence cubes for non-native taxa in Belgium and Europe. 20200204. Zenodo. Release date: 2020-4-02. URL: https://doi.org/10.5281/zenodo.3635510
- Oldoni D, Groom Q, Adriaens T, Davis AS, Reyserhove L, Strubbe D, Vanderhoeven S, Desmet P (2020b) Occurrence cubes at species level for European countries. 20200205. Zenodo. Release date: 2020-2-05. URL: https://zenodo.org/record/3637911
- Vanderhoeven S, Adriaens T, Desmet P, Strubbe D, Backeljau T, Barbier Y, Brosens D, Cigar J, Coupremanne M, De Troch R, Eggermont H, Heughebaert A, Hostens K,

Huybrechts P, Jacquemart A, Lens L, Monty A, Paquet J, Prévot C, Robertson T, Termonia P, Van De Kerchove R, Van Hoey G, Van Schaeybroeck B, Vercayie D, Verleye T, Welby S, Groom Q (2017) Tracking Invasive Alien Species (TrIAS): Building a data-driven framework to inform policy. Research Ideas and Outcomes 3 https://doi.org/10.3897/rio.3.e13414