ECOGRAPHY

Software notes

occCite: Tools for querying and managing large biodiversity occurrence datasets

Hannah L. Owens, Cory Merow, Brian S. Maitner, Jamie M. Kass, Vijay Barve and Robert P. Guralnick

Hannah L. Owens (https://orcid.org/0000-0003-0071-1745) ✉ (hannah.owens@sund.ku.dk), Center for Macroecology, Evolution, and Climate, GLOBE Inst., Univ. of Copenhagen, Denmark; Florida Museum of Natural History, Univ. of Florida, USA. – Cory Merow, Eversource Energy Center, Univ. of Connecticut, Storrs, USA; Dept of Ecology and Evolutionary Biology, Univ. of Connecticut, Storrs, USA. – Brian Maitner (https://orcid.org/0000-0002-2118-9880), Dept of Ecology and Evolutionary Biology, Univ. of Connecticut, USA. – Jamie M. Kass (https://orcid.org/0000-0002-9432-895X), Biodiversity and Biocomplexity Unit, Okinawa Inst. of Science and Technology Graduate Univ., Onna-son, Kunigami-gun, Okinawa, Japan. – Vijay Barve (https://orcid.org/0000-0002-4852-2567), Dept of Entomology, Purdue Univ., Indiana, USA; Florida Museum of Natural History, Univ. of Florida, USA. – Robert P. Guralnick (https://orcid.org/0000-0001-6682-1504), Florida Museum of Natural History, Univ. of Florida, USA.

Ecography
44: 1–8, 2021
doi: 10.1111/ecog.05618

Subject Editor: Thiago F. Rangel
Editor-in-Chief: Miguel Araújo
Accepted 7 May 2021

The amount of observational and specimen-based biodiversity data available to researchers is increasing exponentially, yet the ability to manage and cite large, complex biodiversity datasets lags behind. This management and citation gap impedes reproducibility for data users and the ability for data publishers to track use and accumulate use citations, ultimately harming the longer-term sustainability of the still-emerging enterprise of research data-sharing. Here we present an R package, *occCite* (v. 0.4.7), to aid researchers in querying large species occurrence data aggregators (specifically, the Global Biodiversity Information Facility, GBIF, and the Botanical Information and Ecology Network, BIEN), and store metadata such as primary data providers, database accession dates, DOIs, and the taxonomic source used for search terms. *occCite* also includes tools to summarize and visualize query results and generate citation lists of all data providers and software packages used during the query process. We provide examples of a basic occurrence search and citation workflow as well as an advanced workflow using features for custom optimized searches, visualization, and summary procedures. *occCite* improves upon existing R packages by uniting data from powerful API-based query packages (*rgbif* and *BIEN*) into a unified object-based framework, while maintaining metadata vital to best-practice recommendations for documenting biodiversity analysis workflows. *occCite* aims to efficiently close the gap in the citation cycle between primary data providers and final research products, allowing researchers to meet dataset documentation standards without sacrificing time and resources to the demands of providing increasing levels of detail on their datasets.

Keywords: citations, database aggregation, metadata, presence-only data, R package

NORDIC SOCIETY OIKOS

www.ecography.org

Background

Recent advances in standards development (e.g. the Darwin Core; Wieczorek et al. 2012) and data publication methods (Robertson et al. 2014) have catalyzed cloud-based, open sharing of digitized natural history data in consistent formats (Constable et al. 2010). The result is over 1.6 billion occurrence records for species from across the tree of life served by the Global Biodiversity Information Facility (GBIF) alone (GBIF Secretariat 2020). As of November 2020, over 5000 peer-reviewed journal articles have been published that cite these data (GBIF Secretariat 2020), on topics ranging from biodiversity and biogeography to agriculture and climate change (Ball-Damerow et al. 2019).

These digitally accessible data are the result of millions of cumulative person-hours spent not only collecting organisms in the field, but also preparing and accessioning observations, specimens, and metadata post-collection (Hedrick et al. 2020). These data are also constantly being updated and revised as taxonomy changes and more specimens are accrued and digitized. Therefore, accession dates are a key piece of metadata to identify and trace possible data issues (Feng et al. 2019). However, despite an increasing interest in formalizing standards and metadata protocols for biodiversity data (Feng et al. 2019, Merow et al. 2019, Zurell et al. 2020), tools to manage the connections between occurrence data and the providers of those data, as well as assure proper citation as a key part of this process, are still nascent. Following best practices for citing datasets in a way that ensures a study is truly repeatable remains challenging given the size, complexity, and dynamism inherent in aggregating natural history data.

We must preserve the cycle of data citation from primary data sources to aggregating databases to research products and back again to primary data sources (Escribano et al. 2018). The citation cycle facilitates reproducibility and scientific transparency, but it is also key to supporting primary providers by documenting the use of their data. Data aggregators such as the Global Biodiversity Information Facility (GBIF) have made great strides in harvesting citations from research products and linking them back to primary data providers (Noesgaard 2019). However, this cycle functions only if those who publish research products cite primary data sources – in 2018, only 15% of research studies using GBIF data included the digital object identifier (DOI) for datasets provided by GBIF (Noesgaard 2019). Further, the R statistical computing environment (<https://www.R-project.org/>), which is used heavily by biodiversity researchers, has surprisingly few mechanisms to facilitate proper citation, unlike access through the GBIF web portal. In an era of shrinking funding for natural history museums and community science initiatives, but increasing relevance of the biodiversity they document, it is important to cite these primary data providers to highlight their efforts and emphasize their role as an essential link in the research chain.

Some packages in R provide tools that enable researchers to document sources during the data collection process. For example, *rgbif* (Chamberlain et al. 2020a) for GBIF and *BIEN* (Maitner 2020) for the Botanical Information and Ecology Network (BIEN) provide interfaces for their specific aggregator databases that include valuable features for citing data. However, these and other R packages that serve a single aggregator database are designed for specific use cases tailored to their databases, and uniting aggregator results into a single dataset brings its own set of challenges. Multi-platform occurrence aggregators do exist – searches using *spocc* (Chamberlain 2019) can return occurrence information from up to six aggregator databases – but the process of combining data from these aggregators in each query results in the loss of key metadata: particularly accession date, primary data source, and in the case of GBIF, dataset DOIs. This is particularly important for software that uses occurrence downloading tools to supply data for biodiversity analyses (Kass et al. 2018, Osorio-Olvera et al. 2020). Finally, to our knowledge, there remains a deficit of R tools that manage metadata from an occurrence search and translate it into citations for primary data providers that include accession dates.

We here present a new package, *occCite*, which eases the burden on researchers to document increasingly complex occurrence datasets and associated metadata, with the aim of making studies on large, aggregated databases truly repeatable. *occCite* enables users to download data from multiple aggregator databases; manage, summarize, and visualize multi-database search results; and complete the data citation cycle by generating primary provider citations with DOIs and accession dates. *occCite* therefore preserves links between occurrence data and primary providers throughout the data-processing workflow. This package was initially developed as a module for dataset citation within the *Wallace* ecological modeling application (Kass et al. 2018) but was engineered to also have standalone functionality. We hope this package will enable more studies to reach emerging standards (Feng et al. 2019, Merow et al. 2019, Zurell et al. 2020) for occurrence citation practices that are fully open, repeatable, and acknowledge primary data providers for their hard work assembling, digitizing and publishing data.

Package overview

A stable version of *occCite* (ver. 0.4.7) is available via CRAN <https://CRAN.R-project.org/package=occCite>; the package is currently under review for inclusion in ROpenSci <https://ropensci.org/packages/> and the developer version can be accessed via GitHub <https://github.com/hannahlowens/occCite>. At its simplest, the *occCite* workflow follows a two-step process (Fig. 1). First, the user enters the names of one or more taxa into *occQuery()* and optionally, their GBIF login information (registration is free via the GBIF website and is required to access full metadata; <www.gbif.org>); *occCite* then checks these taxon names and searches for occurrence data via queries to the BIEN database (through *BIEN*) and/or the GBIF database (through *rgbif*). Search results and metadata are contained

in an *occCiteData* object, both in their raw form and as a single table of results for each species with the date of observation, latitude, longitude, primary data source, and database aggregator source. The raw search returns are written to local memory for metadata purposes and to allow the user to access fields other than those in the single processed results table. The user can then pass the *occCiteData* object to *occCitation()*, which compiles citations and accession dates for the primary data providers based on metadata provided by BIEN and/or GBIF. *occCiteData* includes a *summary()* method that returns the taxonomic rectification and/or *occQuery()* search and associated metadata, and *occCitation()* includes a *print()* method that returns a formatted, alphabetized block of text with citations for each primary data provider represented in the search results. These text citations are one potential mechanism for data citation, as part of literature-cited sections (Riemer et al. 2018) or supporting information.

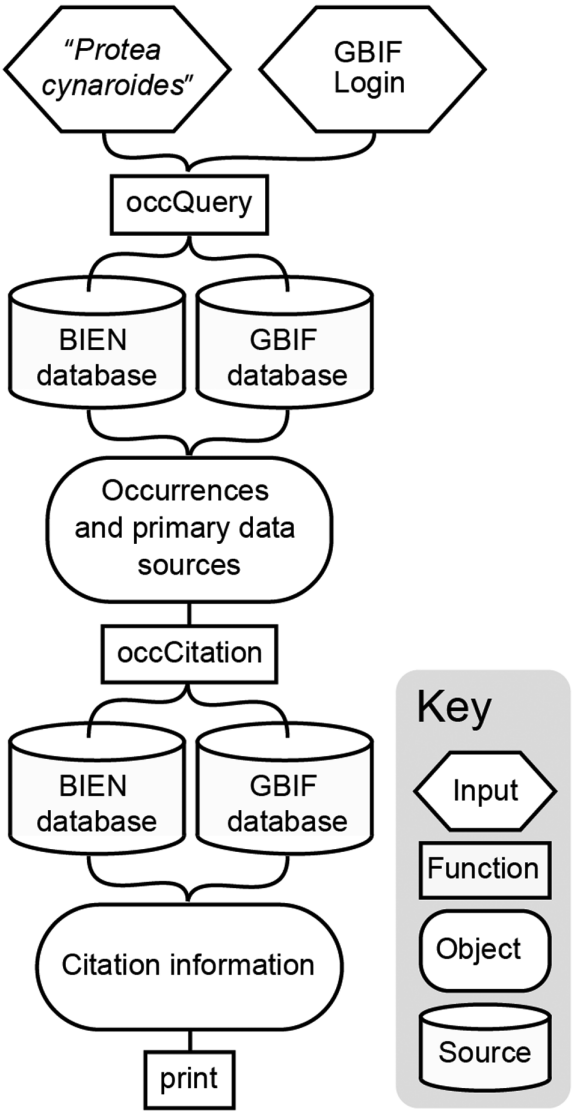

Figure 1. Schematic of basic *occCite* workflow.

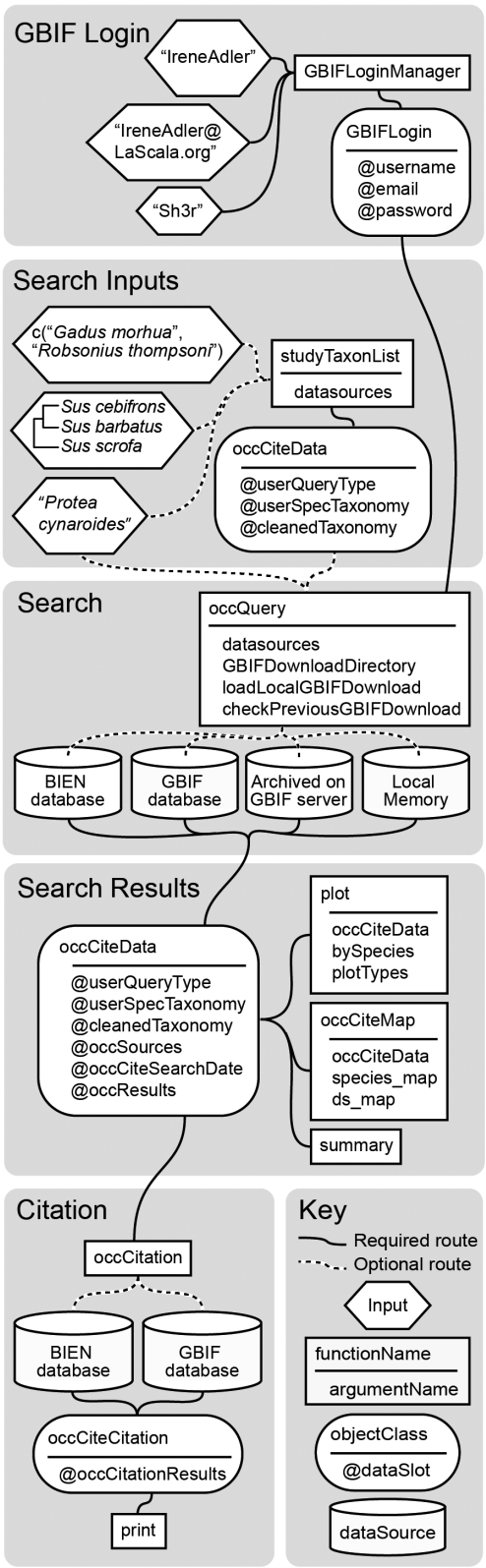

Figure 2. Schematic of full *occCite* architecture, including optional and required workflow routes.

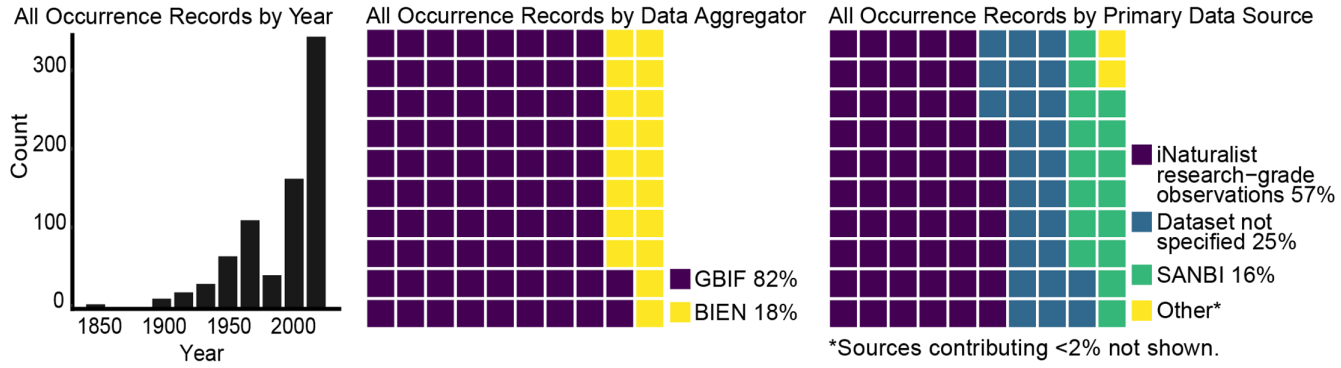

Figure 3. Results of *Protea cynaroides* query from Example 1 visualized using the *sumFig()* function.

Package details

While the basic *occCite* workflow is quite simple, we have designed several options and features (Fig. 2) to allow users to build more customized workflows and visualize the results of their searches both as graphs (Fig. 3) and interactive maps generated using the *leaflet* package (Cheng et al. 2019). What follows is a detailed explanation of the architecture and options available to *occCite* users to optimize the workflow to their specific needs. Novel *occCite* functions, methods, and objects are bolded and italicized. A supporting information vignette that demonstrates these package details and includes examples of citation output is available <https://hannahlowens.github.io/occCite/>.

Setup

We provided a dummy login in Example 1 and 2 to illustrate the format. A login is required because *occQuery()* is, in part, a wrapper around *occ_download()* from the *rgbif* package – this function is analogous to requesting a

Example 1. A single-species search. Refer to package documentation (*?occCite* in R) for function and argument details.

```
library(occCite)

# Setup ----
# Creating a GBIF login
GBIFLogin <- GBIFLoginManager(user = "occCiteTester",
                              email = "****@yahoo.com",
                              pwd = "12345")

# Simple search ----
simpleOC <- occQuery(x = "Protea cynaroides",
                     datasources = c("gbif", "bien"),
                     GBIFLogin = GBIFLogin,
                     GBIFDownloadDirectory = "USER/DIR/",
                     checkPreviousGBIFDownload = FALSE)

# Using the summary function
summary(simpleOC)

# Get citations ----
simpleOccCitations <- occCitation(simpleOC)
print(simpleOccCitations)

# Visualization features ----
plot(simpleOC,
     bySpecies = FALSE,
     plotTypes = c("yearHistogram", "source", "aggregator"))
occCiteMap(simpleOC, cluster = TRUE)
```

Example 2. Loading previously downloaded data into *occCite* from a query based on a phylogeny.

```
library(ape)
library(occCite)

# Setup ----
# Creating a GBIF login
GBIFLogin <- GBIFLoginManager(user = "occCiteTester",
                              email = "****@yahoo.com",
                              pwd = "12345")

# Search for occurrences using a phylogeny
rawTree <- system.file("extdata/Fish_12Tax_time_calibrated.tre",
                       package="occCite")
tree <- read.nexus(rawTree)

# Query databases for names
phyOC <- studyTaxonList(x = tree, datasources = "NCBI")

# Load downloaded GBIF data from local machine ----
phyOC <- occQuery(x = phyOC,
                  datasources = "gbif",
                  GBIFDownloadDirectory = system.file("extdata",
                                                      package = "occCite"),
                  loadLocalGBIFDownload = TRUE,
                  checkPreviousGBIFDownload = FALSE)
summary(phyOC)

# Citations ----
phyOccCitations <- occCitation(phyOC)

# Print citations by species
print(phyOccCitations, bySpecies = TRUE)

# Visualization features ----
# Generate summary figures by species
plot(phyOC,
     bySpecies = TRUE,
     plotTypes = c("yearHistogram", "source"))

# Mapping select species occurrences from full dataset
occCiteMap(phyOC,
           species_map = "Kajikia albida",
           species_colors = "red")
```

doi-referenced dataset download via the GBIF website (Chamberlain et al. 2020). The username, email, and password are stored in the R working environment as a *GBIFLogin* object when they are supplied to *occCite*'s *GBIFLoginManager()* function to simplify their specification for users.

Taxonomic rectification

In the simplest of searches, such as in Example 1, the input species' name is automatically checked for spelling errors and taxonomic validity through the *occQuery()* function using the National Center for Biotechnology Information (NCBI) taxonomy <https://www.ncbi.nlm.nih.gov/taxonomy>. This is done via *gnr_resolve()* from the *taxize* R package (Chamberlain et al. 2020b) – this function automatically checks spelling and presence of the name in the NCBI taxonomy (the 'cleanedTaxonomy' slot of the *occCite* object contains a table with the user's name and the name from the taxonomy that best matches the input name). The *studyTaxonList()* function is provided for better control over taxonomic rectification and will accept either a vector of species' names or a phylogeny of class *phylo* (Example 2). Using

this function, it is possible for the user to choose from any of the available taxonomies in the Global Names Index <http://gni.globalnames.org/>, a component of the Global Names Architecture (Patterson et al. 2010). *studyTaxonList()* creates an *occCiteData* object, which can then be passed into *occQuery()* to perform an occurrence data search. Names without a match in the taxonomy of choice are returned via warning messaging and flagged in the *occCiteData* 'cleanedTaxonomy' slot to facilitate user review. Currently, *occQuery()* searches only for occurrence data that matches the input name in the occurrence database of choice, and thus does not return records corresponding to synonyms, misspellings, and other errors in the database(s).

Query

The *occQuery()* function is designed to provide users with several ways to generate and optimize repeatable occurrence searches while keeping detailed metadata. *occQuery()* returns an *occCiteData* object that stores information on the type of query made (i.e. user-supplied list or phylogeny), the date of the query, the taxonomic resources used for name rectification, the accepted taxonomic names used in the search, the database aggregators searched, and a named list of search results corresponding to the taxonomic names used in the search (Fig. 2). There are also several optional arguments for *occQuery()* to load local downloads of GBIF data as well as previously prepared downloads being stored on GBIF's servers; these arguments are detailed below under 'Advanced Features'.

Citations

After the occurrence data search is complete, the resulting *occCiteData* object can be passed to *occCitation()* to generate citations for primary biodiversity databases. *occCitation()* returns an *occCiteCitation()* data object, which is a named list with entries corresponding to the taxonomic names used to build a query. Each item in the list is a data frame with one row for each primary data provider to be cited. Columns include the name of the database aggregator, unique identifier code for the primary provider record as used by the database aggregator, the citation and accession date for the primary provider, and the number of occurrences supplied by that data provider. The *print()* method for an *occCiteCitation* object returns a formatted and alphabetized set of references, either as a single block of text for all species appropriate for addition to the references section of a publication, or as separate blocks of text for each species individually for more detailed source-parsing in publication Supporting information or other documentation. Examples of these outputs can be found in the package vignette <https://hannahlowens.github.io/occCite/>.

Advanced features

Downloading data from GBIF can be time-consuming, especially for multiple species and/or species with many occurrence records. To save time when repeating a query that has been run in the past, the user has two options: 1) download previously prepared datasets from the GBIF servers (stored for six months after initial download request; GBIF Secretariat 2020); or 2) access previously downloaded datasets stored on their local machine. By default, *occQuery()* checks GBIF's servers for the user's previously prepared datasets before preparing a new dataset (this behavior can be disabled by setting the *checkPreviousGBIFDownload* argument to 'FALSE'). Alternatively, if the user wishes to access downloaded GBIF dataset *.zip* files on their local machine, the *loadLocalGBIFDownload* argument must be changed to 'TRUE' and the directory where the files are located must be specified via the *GBIFDownloadDirectory* argument. *occQuery()* will crawl through the specified directory and collect all the downloaded datasets contained in that folder and its subfolders. It will then import the most recent downloads for each species in the taxon list into the R working environment. These GBIF data can then be appended to a BIEN search (if desired) in the same way as if the user conducted a simple real-time search (Example 1); acquiring citation data follows the same set of steps as Example 1 (Example 2). *occCite* does not currently support mixed data download sources—that is, it is not possible to download GBIF datasets for some species and load the rest from local .zip files.

Discussion

As the literature on biodiversity modeling expands, difficulties associated with generating and managing appropriate metadata to render these studies reproducible will continue to grow. Three recent papers have outlined complementary and interconnected visions for biodiversity model metadata reporting standards (Feng et al. 2019, Merow et al. 2019, Zurell et al. 2020). All three agree that sources of occurrence data are a basic necessity of model documentation workflows, although they differ in recommendations regarding the necessary level of detail for these data. Feng and colleagues (2019) reviewed recent literature and generated a checklist for reporting on modeling methods, recommending that researchers report the source of their occurrence data, as well as the download date and/or version of the data source used. Merow and colleagues (2019) did not make such a specific recommendation, but their Range Model Metadata Standards (RMMS) framework does require occurrence data sources to be reported. Most recently, Zurell and colleagues (2020) expanded on the RMMS data dictionary for their Overview, Data, Model, Assessment, and Prediction (ODMAP) protocol, designed specifically for species distribution model studies. They recommend not only that occurrence data sources be cited with accession dates, but that sample size per taxon and taxonomic reference system be reported.

Both RMMS and ODMAP provide tools to generate a metadata document that supplement more traditional methods sections in biodiversity studies. However, neither of these sets of tools is designed to directly manage the complex stream of occurrence data upon which many biodiversity studies are

built. *occCite* closes this gap by managing occurrence query information including dataset aggregators, accession dates, and taxonomic reference systems, as well as the primary data sources, observation dates, and per-taxon sample sizes of resulting occurrence records, along with reporting methods that can be directly fed into the RMMS workflow or copied into ODMAP.

occCite's utility in the context of biodiversity studies is clear, but its potential applications extend beyond facilitating citations for range modeling. We designed *occCite* to accept phylogenies into occurrence queries as a first step towards building documentation protocols for spatial comparative phylogenetics. By combining information on the evolutionary relationships among taxa with data on where those taxa are found, we can begin to more fully understand how ecological, geological, and evolutionary processes have shaped past and present biodiversity patterns. These inferences can then provide insight into the distributions of biodiversity in the future (Quintero and Wiens 2013, Jezkova and Wiens 2016).

The sheer amount of biodiversity data is growing every day, but documenting its use in scientific research following best-practice standards (Feng et al. 2019, Merow et al. 2019, Zurell et al. 2020) has not kept pace. As datasets and analyses continue to grow in size and complexity, ensuring acquisition and analysis protocols are completely reproducible requires increasing time and resources. Furthermore, in the process of aggregating data from many sources, citation linkages to primary data providers can be unintentionally severed as workflows and tools in common use are not well developed for this key data management task. *occCite* was built to keep all the ease of using existing tools but with the goal of significantly simplifying data citation production and improving reproducibility, as users are able to also more easily manage data resources stored either locally or on a cloud server. *occCite* also already integrates with the ecological modeling platform *Wallace* (Kass et al. 2018), thus enhancing existing, well-used tools meant to enhance best-practice species distribution modeling and data management frameworks.

Acknowledgements – We thank members of the Guralnick, Soltis, and Kawahara labs at the University of Florida for their contributions to beta-testing the *occCite* software, and members of the *Wallace* development team in the Anderson lab at the City College of New York, CUNY for additional discussion and feedback.
Funding – Funding for this project was provided by a seed grant from the University of Florida Biodiversity and Informatics Institutes and a second place Ebbe Nielsen Challenge prize from the Global Biodiversity Information Facility. CM acknowledges funding from NSF grant DBI-1913673 and DBI-1661510.

Author contributions

Hannah L. Owens: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Funding acquisition (equal); Investigation (lead); Methodology (lead); Project administration (lead); Resources (equal); Software (lead); Visualization (lead); Validation (lead); Writing – original draft (lead); Writing – review and editing (lead). **Cory Merow**: Conceptualization (supporting); Formal analysis (supporting); Funding acquisition (equal); Investigation (equal); Methodology (supporting); Software (supporting); Validation (supporting); Visualization (supporting); Writing – review and editing (equal). **Brian S. Maitner**: Conceptualization (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Software (supporting); Validation (supporting); Visualization (supporting); Writing – review and editing (supporting). **Jamie M. Kass**: Conceptualization (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Software (supporting); Validation (supporting); Visualization (supporting); Writing – review and editing (supporting). **Vijay Barve**: Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Software (supporting); Validation (supporting); Writing – review and editing (supporting). **Robert P. Guralnick**: Conceptualization (equal); Data curation (supporting); Formal analysis (supporting); Funding acquisition (equal); Investigation (supporting); Methodology (supporting); Project administration (supporting); Resources (supporting); Software (supporting); Supervision (supporting); Validation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review and editing (supporting).

Transparent Peer Review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.05618>.

Data availability statement

A vignette and further documentation for the *occCite* package are available at the package website: <https://hannahlowens.github.io/occCite/>. CRAN release: <https://cran.r-project.org/package=occCite>; ROpenSci release: under review; GitHub development page: <https://github.com/hannahlowens/occCite>.

To cite *occCite* or acknowledge its use, cite this Software Note as follows, substituting the version of the application that you used for 'version 0':

Owens, H.L., Merow, C.B., Maitner, B.S., Kass, J.M., Barve, V.V., and Guralnick, R.P. 2021. occCite: Tools for Querying and Managing Large Biodiversity Occurrence Datasets. Ecography 44: XXX–XXX (ver. 0).

References

Ball-Damerow, J. E. et al. 2019. Research applications of primary biodiversity databases in the digital age. – PloS One 14: e0215794 .
Chamberlain, S. 2019. spocc: Interface to species occurrence data sources ver. 1.0.2. – <https://CRAN.R-project.org/package=spocc>.

Chamberlain, S. et al. 2020a. rgbif: Interface to the Global Biodiversity Information Facility API ver. 3.2.0. – <http://doi.org/10.5281/zenodo.3956837>.
Chamberlain, S. et al. 2020b. taxize: Taxonomic information from around the web version 0.9.95. – <http://doi.org/10.5281/zenodo.7097>.
Cheng, J. et al. 2019. Leaflet: create interactive web maps with the JavaScript 'Leaflet' library ver. 2.0.3. – <https://CRAN.R-project.org/package=leaflet>.
Escribano, N. et al. 2018. The tragedy of the biodiversity data commons: a data impediment creeping nigher? – <https://doi.org/10.1093/database/bay033>.
Feng, X. et al. 2019. A checklist for maximizing reproducibility of ecological niche models. – Nat. Ecol. Evol. 3: 1382–1395.
GBIF Secretariat 2020. GBIF: Global Biodiversity Information Facility. – <www.gbif.org>, accessed 21 September 2020.
Hedrick, B. P. et al. 2020. Digitization and the future of natural history collections. – BioScience 70: 243–251.
Jezkova, T. and Wiens, J.J. 2016. Rates of change in climatic niches in plant and animal populations are much slower than projected climate change. – P. R. Soc. B. 283: 28320162104, <https://doi.org/10.1098/rspb.2016.2104>.
Kass, J. M. et al. 2018. Wallace: a flexible platform for reproducible modeling of species niches and distributions built for community expansion. – Methods Ecol. Evol. 9: 1151–1156.
Maitner, B. 2020. BIEN: Tools for accessing the botanical information and ecology network database ver. 1.2.4. – <https://CRAN.R-project.org/package=BIEN>.
Merow, C. et al. 2019. Species' range model metadata standards: RMMS. – Global Ecol. Biogeogr. 28: 1912–1924.
Noesgaard D. 2019. Improving impact metrics of open and free biodiversity data through LinkedMetadata and Academic outreach. – Biodivers. Inf. Sci. Stand. 3: e35723.
Osorio-Olvera, L. et al. 2020. ntbox: An r package with graphical user interface for modelling and evaluating multidimensional ecological niches. – Methods Ecol. Evol. 11: 1199–1206.
Patterson, D. J. et al. 2010. Names are key to the big new biology. – Trends Ecol. Evol. 25: 686–691.
Quintero I. and Wiens J. J. 2013. Rates of projected climate change dramatically exceed past rates of climatic niche evolution among vertebrate species. – Ecol. Lett. 16: 1095–1103.
Riemer, K. et al. 2018. No general relationship between mass and temperature in endothermic species. – Elife 7: e27166.
Robertson, T. et al. 2014. The GBIF integrated publishing toolkit: facilitating the efficient publishing of biodiversity data on the internet. – PloS One 9: e102623.
Wieczorek J. et al. 2012. Darwin core: an evolving community-developed biodiversity data standard. – PloS One 7: e29715.
Zurell, D. et al. 2020. A standard protocol for reporting species distribution models. – Ecography 3: 1261–1277.