

## Joe Millard—Data scientist and computational ecologist

Address: 52 Aits View, Victoria Avenue, West Molesey, Surrey, KT81TL

Email: joeymillard@hotmail.co.uk; Mobile: 0758 4168 355; Website: <https://joemillard.github.io/>;

GitHub: <https://github.com/Joemillard>

### Summary

I'm a postdoctoral researcher at the Natural History Museum (London), primarily interested in the causes and consequences of global pollinator biodiversity change, and the power of monitoring biodiversity change and public biodiversity awareness in real-time. Prior to my current postdoc I worked at the Leverhulme Centre for Demographic Science as a data scientist and researcher. I completed a PhD in computational ecology jointly between UCL, the Institute of Zoology, and the RSPB. Previously I also worked at UNEP-WCMC building automated data handling systems for wildlife trade data, and at the journal Scientific Reports in manuscript peer-review and decisions. I'm also an award-winning science writer and R Shiny developer.

### Education and qualifications

#### **PhD Computational Ecology (Minor corrections) – UCL London NERC DTP (2017-2021)**

Thesis: Causes and consequences of global pollinator biodiversity change in the Anthropocene

- Undertook a CASE (Collaborative Award in Science and Engineering) PhD in collaboration with UCL (Dr Tim Newbold), the Institute of Zoology (Dr Robin Freeman), and the RSPB (Prof Richard Gregory).
- My project aimed to collate current knowledge on pollinator decline, before building statistical models to predict the response of pollinators to future environmental change. I also explored developments in the novel field of conservation culturomics, using Wikipedia web-scraping to develop metrics for biodiversity awareness. My PhD applied a broad spectrum of big data tools in R and Python.
- I published 3 first author papers in high-quality peer-reviewed journals (*Nature Communications*, *Conservation Biology*, and *Ecography*).
- My paper in *Conservation Biology* appeared in a special section on 'Advancing Conservation Culturomics' research, and my paper in *Nature Communications* was selected for an 'Editor's Highlight'.

#### **MSc Biodiversity and Conservation (Distinction, 1<sup>st</sup>) – The University of Leeds (2014-2015)**

Thesis: A Satellite Telemetry Investigation of Caspian Seal (*Pusa caspica*) Activity Budgets (77)

- Awarded the second highest overall mark in the year for my data science research project and the highest thesis mark at 80.5%.
- Gained experience in large-scale data analysis using R packages ggplot2, reshape2, lme4, ggmap, and gganimate.

#### **BSc Zoology, Upper second-class honours (2:1) – The University of Leeds (2011-2014)**

Thesis: Of Molecules and Morphology: An Extended Review of the Eutherian Radiation (69)

### Employment and volunteering summary

#### **Postdoctoral researcher – the Natural History Museum, London (March 2022-present)**

- Currently funded on the NERC project GLiTRS, a cross-institutional consortium attempting to build a global threat-response model for insect biodiversity change.
- Lead researcher in combining multiple data types (expert elicitation, space-for-time, and meta-analyses) to develop the core threat-response model for insect biodiversity.
- Supervised 6 masters students (primary supervisor, 5; co-supervisor, 2), led catch-up meetings with groups of masters students and post docs, introduced new starter students to the museum, and NHM lead on GLiTRS recruitment during line manager absence.

**Postdoctoral data scientist – Leverhulme Centre for Demographic Science & Nuffield College, University of Oxford** (September 2021-March 2022)

- Funded by Prof Melinda Mills on the ERC Advanced Grant CHRONO, an interdisciplinary project on sleep and chronotype disruption.
- Lead on three research projects: 1) a systematic text-analysis review of the causes of human sleep disturbance; and 2) changes in Twitter activity following the COVID-19 pandemic; and 3) the effects of extreme night-time temperature on sleep disturbance, using ERA-5 hourly temperature data and Twitter scraping.
- Helped design a set of data management plans for the collection of multiple forms of publicly donated data (e.g. social media, genetic, genealogy, GPS), and worked with external companies to ensure a secure backend database.
- Co-authored a SAGE (SPI-B) Number 10 cabinet briefing on the reintroduction of COVID-19 preventative measures following the Autumn/Winter 2021 plan.

**Postgraduate teaching assistant** (2019-2021)

- Taught statistical programming on two computational biology modules, for a total of more than 100 hours. Covered statistical methods (lme4, anova, lm), programming (basic Python), data handling (piping, dplyr verbs), and functional programming in R.

**Programme Assistant – UNEP-WCMC** (September 2016-June 2017)

- Collaborated with Dr Aly Pavitt to develop two new automated systems in R for the processing of CITES data, replacing a previous system built in VBA, and bringing processing time down from 3-4 hours to 5-10 seconds.
- Successfully pitched to Aly that we should implement data cleaning scripts as R Shiny apps, enabling non-technical team members to benefit from the greater speed of R over VBA.
- Built the front end Shiny app, combined the UI and server side scripts, tested the app in the initial phase, and wrote a manual.
- Delivered workshops on Shiny to the Species Programme, the wider centre, and a specialist technical GIS hub.
- Involved in producing a data handling guide for the Ecuadorian environmental ministry and an Amazonian regional trade overview.
- Designed a new interactive data visualisation CITES tool for the WCMC informatics team.
- Wrote ~12 high quality wildlife trade reviews for the European Commission.

**Publishing Assistant – Nature Publishing Group** (November 2015-September 2016)

- Applied my broad academic knowledge to rapidly assign manuscripts to the editors at Scientific Reports, and then responsible for manuscripts to decision.
- Trained in the handling and identification of clinical trials and responsible for vetting trials before acceptance.
- Responsible for the growth and maintenance of the Ecology and Evolution section of the editorial board.
- Presented accepted ecological papers to the editorial team and wrote paper summaries for our Facebook page, published to more than 100,000 followers.

**Academic tutor – UKonexão** (2018-2019)

- Set up a private coaching service for Brazilian students wanting to study in the UK.
- Delivered nine 1 hour 1-1 Skype advice sessions for students at universities in Brazil.
- Provided verbal and written comments on CVs for 9 undergraduate and postgraduate students in Brazil, and gave a talk in person at the University of São Paulo on studying in the UK.

**Social Media Coordinator and Blogger – Eco Sapien (November 2014–November 2015) & Graph Start (September 2015–September 2018)**

- Key team member of Eco Sapien, a non-profit organisation dedicated to ecological education.
- Published more than 50 popular blog posts and 15 extended articles, and occasionally gained over 1000 reblogs and likes for my posts.
- Set up Graph Start, an educational website dedicated to data visualisation in R.
- Wrote blogs on R programming and took on three voluntary team members: one graphic designer to assist with branding and two bio-statisticians to provide R tutorials.

**Lead Conference Coordinator and Speaker (January 2015)**

- Won a place as lead coordinator on the first IEMA Student Sustainability Conference, a national event with more than 100 delegates.
- Delivered an introductory speech, ran two conference sessions, devised timetables, booked caterers, researched marketing strategies, and later invited to meet the CEO of IEMA.

**Research Assistant – Action for the Wild (Summer 2014) & Cetacean Research and Rescue Unit (June 2016)**

- Worked as a field assistant for 2 months on the Umphafa Private Nature Reserve, South Africa, and for 2 weeks in the Moray Firth.
- Gained core field research skills and an ability to lead and assist undergraduate researchers.

**Peer-reviewed publications (3 first author, 2 group lead last author)**

1. Newbold, T., Adams, G.L., Robles, G.A., Boakes, E.H., Ferreira, G.B., Chapman, A.S., Etard, A., Gibb, R., Outhwaite, C.L., **Millard, J.**, and Williams, J.J., 2019. Climate and land-use change homogenise terrestrial biodiversity, with consequences for ecosystem functioning and human well-being. *Emerging Topics in Life Sciences*. DOI: <https://doi.org/10.1042/ETLS20180135>
2. **Millard, J.**, Freeman, R., Newbold, T., 2020. Text-analysis reveals taxonomic and geographic disparities in animal pollination literature. *Ecography*. DOI: <https://doi.org/10.1111/cobi.13701>
3. **Millard, J.**, Gregory, R., Jones, K., Freeman, R., 2021. The Species Awareness Index: a Wikipedia derived conservation culturomics metric for public biodiversity awareness. *Conservation Biology* (selected to appear in an 'Advancing Conservation Culturomics' special section). DOI: <https://doi.org/10.1111/cobi.13701>
4. **Millard, J.**, Outhwaite, C., Kinnersley, R., Freeman, R., Gregory, R., Adedaja, O., Gavini, S., Kioko, E., Kuhlmann, M., Ollerton, J., Ren, Z-X., and Newbold, T., 2021. Global effects of land-use intensity on local pollinator biodiversity. *Nature Communications* (selected as an Editor's Highlight). DOI: <https://doi.org/10.1038/s41467-021-23228-3>
5. Cornford, R., **Millard, J.**, Gonzalez-Suarez, M., Freeman, R., Johnson, T. F., Automated synthesis of biodiversity knowledge requires better tools and standardised research output. *Ecography* (selected as Editor's Choice). DOI: <https://doi.org/10.1111/ecog.06068>
6. Johnson, T. F., Cornford, R., Dove, S., Freeman, R., **Millard, J.** Achieving a real-time online monitoring system for conservation culturomics. *Conservation Biology*. DOI: <https://doi.org/10.1111/cobi.14096>
7. Skinner, G., Cooke, R., Keum, J., Purvis, A., Raw, C., Woodcock, B., **Millard, J.** Dynameta: a dynamic platform for ecological meta-analyses in R Shiny. *SoftwareX*

**In-prep publications (2 first author)**

1. **Millard, J.**, Akimova, E., Ding, X., Leasure, D., Zhao, B., Mills, M., Stringent COVID-19 government restrictions were associated with a marked increase in Twitter activity in Europe
2. **Millard, J.**, Outhwaite, O., Dicks, L., Ollerton, J., Ceausu, S., Newbold, T., Key tropical crops at risk from pollinator biodiversity loss due to climate change and land use

## Other publications

1. August, Tom A, Affouard, Antoine, Fox, Nathan, Marlowe, Celia, **Millard, J.**, Sanderson, Roy, & Bonnet, Pierre. 2019. AI validated plant observations from social media: Flickr images from central London 2011-2019 (Version 1.0) [Data set]. Zenodo. DOI: <http://doi.org/10.5281/zenodo.3497314>
2. **Millard, J.**, Wikipedia page views reveal global changes in public biodiversity awareness. RSPB guest blog. Link: <https://community.rspb.org.uk/ourwork/b/science/posts/wikipedia-page-views-reveal-global-changes-in-public-biodiversity-awareness>

## Key skills, other information, and hobbies and interests

- **Core data science skills:** Data visualisation (shiny, ggplot2, plotly, gganimate, ggiraph, DiagrammeR, patchwork); data handling (dplyr, reshape2, pandas, numpy); GIS (ggmap, raster, ArcGIS, rgdal, leaflet); database management (MySQL, SQLite); web development and web-scraping (css, html, jquery, WordPress, R Markdown, rvest); statistical programming (linear modelling, mixed-effects modelling); natural language processing (stringr, taxize, named-entity recognition, tf-idf, cosine similarity); geoparsing (CLIFF-CLAVIN); version control and reproducibility (checkpoint, pacman, git, GitHub, here::here); parallel computing (HPC, parLapply).
- **Other information:** Full, clean driving license; RLSS Bronze Medallion; languages (English; Portuguese, pre-intermediate—A2).
- **Hobbies and interests:** Reading (sci-fi, popular science, and history); watching movies (sci-fi, thrillers, documentaries); travelling; outdoor activities (camping, walking, kayaking); eating out and socialising with family and friends.

## Invited talks and Symposia

- **Royal Society Recovering Nature event: building on Georgina Mace's work to ensure a biodiverse and liveable future** – invited as one of 12 contributed speakers building on Georgina's work
- **ICCB 2023 round-table symposium; Large scale monitoring of perceptions of nature: state of the art and ways forward** – co-author on accepted proposal, and then group lead on presentation entitled "Achieving a real-time online monitoring system for conservation culturomics"
- **NHM & Turing AI Lunch Series** – invite only lunch for AI scientists working in the natural and environmental sciences
- **Linnaean Student Conference 2019** – The importance and diversity of animal pollination (<https://www.youtube.com/watch?v=qn0sCxJiTdl&t=5854s>)
- **Planet PhD podcast** – Pollinator biodiversity in a changing world (<https://planetphd.podbean.com/e/joe-millard-pollinator-biodiversity-in-a-changing-world/>)
- **University College London seminar 2019** – Text-analysis reveals taxonomic and geographic disparities in animal pollination literature
- **University of São Paulo, Brazil 2019** – Postgraduate study in the UK: How, when and why?
- **Natural History Museum seminar** – Worldwide vulnerability of local pollinator abundance and crop pollination to land-use and climate change
- **University of Oxford seminar 2021** - Pollinators, culturomics, and chronotype: a summary of my background and ideas for future research
- **Online University College London seminar 2021** – The Species Awareness Index

### Talks and workshops given

- **ENTO22** – Developing a global threat response model for insect biodiversity change
- **IOP Environmental Research 2021** – Worldwide vulnerability of local pollinator abundance and crop pollination to land-use and climate change
- **BES Ecology Across Borders 2021** – Worldwide vulnerability of local pollinator abundance and crop pollination to land-use and climate change
- **BES Festival of Ecology 2020** – The Species Awareness Index: a Wikipedia derived conservation culturomics metric for public biodiversity awareness
- **SCAPE (Scandinavian Association for Pollination Ecology) 2020** – Global effects of land-use intensity on local scale pollinator biodiversity
- **SCCS (Student Conference on Conservation Science) Europe 2019** – Text-analysis reveals taxonomic and geographic disparities in animal pollination literature
- **RSPB Annual Science Meeting 2019** – Text-analysis reveals taxonomic and geographic disparities in animal pollination literature
- **Institute of Zoology 2018 & UCL CBER 2018-2021** – Hosted an IoZ GitHub R clinic and a CBER journal club (fortnightly in-person) and CBER data club (fortnightly online)

### Other conferences, seminars, and workshops attended

- **London R 2018** (Introduction to Data Science) & **Datacamp 2018** (Introduction to SQL)
- **H2O AI conference, London 2019**
- **Nature Research, London 2019** – 2 day manuscript writing workshop
- **Brexit BES early career workshop, London 2019** – 1 day workshop on adapting to Brexit
- **UCL Arena 1 teaching assistant training 2019 & UCL copyright training 2019**
- **UCL intermediate Python programming 2019**

### Prizes, awards, and grants

- **Capgemini Tech4PositiveFutures challenge** – finished as one of the top 6 pitches for a €100k prize applying tech to biodiversity problems
- **PhD Conference travel fund** – £500 grant to attend BES Ecology Across Borders 2021
- **BES Quantitative Ecology Hackathon 2020** – awarded “best presentation” for our plant species identifying AI Shiny app (<https://methodsblog.com/2020/03/12/quantitative-sig-hackathon/>)
- **London NERC PhD studentship (2017-2021)** – awarded funding of ~£70,000 to complete an environmental science PhD on the London NERC DTP
- **Linnaean Student Conference 2019** – won a place as speaker for my abstract on pollinator biology
- **Leeds University Science Magazine Writing competition** – won first prize for my piece “The polar bear: a tale of two futures”
- **IEEMA Student Sustainability Conference** – won a committee place for my essay on sustainability issues