

## Spatial uncertainty in data from Natural History Collections as mediated by GBIF. Implications for ecological research.

Dr. Arnald Marcer<sup>1,2</sup>, Mr. Arthur Chapman<sup>3</sup>, Dr. John Wieczorek<sup>4</sup>, Dr. Xavier Picó<sup>5</sup>, Dr. Francesc Uribe<sup>6</sup>, Mr. John Waller<sup>7</sup>, Dr. Arturo Ariño<sup>8</sup>

<sup>1</sup>CREAF, 08193 Cerdanyola del Vallès, Barcelona, Spain, <sup>2</sup>Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Barcelona, Spain, <sup>3</sup>Australian Biodiversity Information Services, Melbourne, Australia, <sup>4</sup>University of California, Berkeley, USA, <sup>5</sup>Estación Biológica de Doñana (EBD), Consejo Superior de Investigaciones Científicas (CSIC), Sevilla, Spain, <sup>6</sup>Museu de Ciències Naturals de Barcelona, Barcelona, Spain, <sup>7</sup>GBIF, Copenhagen, Denmark, <sup>8</sup>Institute for Biodiversity and Environmental Research (BIOMA) and DATAI, Universidad de Navarra, Pamplona, Spain

MO3) Mobilizing specimens for factbased conservation in the face of a global biodiversity crisis, Lecture Theatre 2, Appleton Tower, 11 Crichton Street, EH8 9LE, June 7, 2022, 4:00 PM - 5:30 PM

Natural history collections (NHC) represent an enormous and largely untapped wealth of information on the Earth's biota, made available through GBIF as digital preserved specimen records. Precise knowledge of where the specimens were collected is paramount to rigorous ecological studies, especially in the field of species distribution modelling. Here, we present a first comprehensive analysis of georeferencing quality for all preserved specimen records served by GBIF, and illustrate the impact that coordinate uncertainty may have on predicted potential distributions. We used all GBIF preserved specimen records to analyse the availability of coordinates and associated spatial uncertainty across geography, spatial resolution, taxonomy, publishing institutions and collecting time. We used three plant species across their native ranges in different parts of the World to show the impact of uncertainty on predicted potential distributions. We found that 38% of the 180+ million records provide coordinates only and 18% coordinates and uncertainty. Georeferencing quality is determined more by country of collection and publishing than by taxonomic group. Distinct georeferencing practices are more determinant than implicit characteristics and georeferencing difficulty of specimens. Availability and quality of records contrasts across world regions. Uncertainty values are not normally distributed but peak at very distinct values, which can be traced back to specific regions of the world. Uncertainty leads to a wide spectrum of range sizes when modelling species distributions, potentially affecting conclusions in biogeographical and climate change studies. In summary, the digitised fraction of the world's NHC are far from optimal in terms of georeferencing and quality mainly depends on where the collections are hosted. A collective effort between communities around NHC institutions, ecological research and data infrastructure is needed to bring the data on a par with its importance and relevance for ecological research.

