# Users and uses of UK Natural History Collections

Natural History Museum

How to cite this paper

Hardy H, Livermore L, Kersey P, Norris K, Smith V, Pullar J (2023) Users and uses of UK Natural History Collections – a Summary
https://doi.org/10.5281/zenodo.8403318

A longer paper on this study including further detail on the methodology and findings is available:
Hardy H, Livermore L, Kersey P, Norris K, Smith V (2023) Understanding the users and uses of UK Natural History Collections.
Research Ideas and Outcomes 9: e113378 https://doi.org/10.3897/rio.9.e113378

# Users and uses of UK Natural History Collections - a Summary

UK natural science collections hold more than 137 million items[1]. These are an unrivalled source of data about how the Earth and its natural systems formed over 4.56 billion years. They contain evidence of the changes in the natural world including the impact of human life on our planet over the past few thousand years. The scientific, commercial, and societal benefits of these collections are currently constrained by the limits of physical access, and by highly fragmented digitisation efforts with less than 10% digitally available. Following work with Frontier Economics in 2021[2], which showed potential for £2 billion in benefits to the UK economy from digitising all UK natural science collections, in 2022-23 the Natural History Museum London worked, with analytical support from McKinsey & Company, to understand the impact of what has already been digitised and shared by UK natural science collections, particularly through the Global Biodiversity Information Facility (GBIF). This paper looks at the demand for these data, what they are used for, and how they underpin efficient, effective and impactful research.

## UK natural science collections impact – summary insights (Jan 2023)

**7.6 million specimens**, from **248 territories and countries**, are freely accessible on GBIF from the 12 UK institutions investigated – these represent less than 6% of total UK natural science collections.

**39 billion individual specimen records** from UK institutions have been downloaded from GBIF - **2,710** publications cite these data.

**12%** of the total peer-reviewed journal articles citing GBIF data cite UK natural science collections – these data currently make up **just 0.3%** of total occurrences on GBIF, **meaning they punch 40 times above their weight.**

In 2022, there was a download event of the Natural History Museum's data from GBIF on average every **3 minutes 24 seconds,** and **2.2** publications per day on average cited UK institutions' data.

**>250 publications** on each of the themes of climate change, invasives and conservation cite UK institutions' uploads – research areas that are key to a future in which people and planet thrive.

**~1,200** UK-affiliated researchers are authors of publications that use UK collections data, among **13,000** researchers from at least **160** countries who have cited UK institutions' GBIF data.

In addition to over £5.4 million invested in digitisation by the Natural History Museum alone, **£18m in efficiencies** have been realised just assuming a single physical visit saved per citation, of which £1.4 million can be attributed to UK-affiliated researchers. These savings can be reinvested in additional research. Interviews with researchers show that real savings can be many times higher, particularly when digital collections data are combined with Artificial Intelegence analysis techniques.

---

1 Smith, V.S., Hardy, H., Wainwright, T., Livermore, L., Fraser, N., Horak, J., Aspinall, J. and Howe, M. (2022) Harnessing the power of natural science collections: a blueprint for the UK, pp 1-28. Natural History Museum. https://doi.org/10.5281/zenodo.6472239

2 Popov D, Roychoudhury P, Hardy H, Livermore L, Norris K (2021) The Value of Digitising Natural History Collections. Research Ideas and Outcomes 7: e78844. https://doi.org/10.3897/rio.7.e78844

# The case for digitisation

Recent research published by major biodiversity data aggregator the Global Biodiversity Information Facility (GBIF) and Deloitte, looking at the value of global GBIF data as a whole[3], shows the extraordinary value of digital biodiversity data to research. This found that nearly 50% of users would not have been able to achieve their research outcomes without GBIF mediated data, and another 41% could only have done so with significantly increased time and effort. Over 90% of users linked their use of GBIF-mediated data to advancing the UN Sustainable Development Goals – natural science collections data are not only relevant to understanding Life on Land and Life Below Water, but also contribute to Climate Action and to many other goals including reducing hunger, poverty, and inequality. Overall, this research showed that every €1 invested in GBIF provides €3 in direct user benefits and up to €12 in societal benefits.

NHM's research similarly identified enormous demand for and benefit from UK collections data, even with the low proportion currently available. Our study was conducted between November 2022 and January 2023, drawing primarily on data from GBIF. This examined data from 12 UK institutions currently publishing primarily collections data to GBIF, combining quantitative and qualitative approaches to investigate the characteristics of uploaded data, users and uses of data, and the value created by that usage.

UK natural science collections data add vital historical, geographic and taxonomic breadth to the data available for research, evidenced by their high rates of citation and through the interviews that we conducted with researchers using collections data. UK institutions provide a significantly higher percentage of specimens from South America, Asia, and Africa (11%, 23% and 16%, respectively) than GBIF occurrence data as a whole (4%, 5% and 3%), and making these global data available to communities of origin and wider global users is an important part of addressing some of the historical legacies of collecting. UK collections also contribute greater taxonomic diversity (e.g., representing a higher percentage of botanic specimens).

UK institution data uploaded to GBIF is used to support research in the UK, as well as 160 countries and territories worldwide. The USA is the top contributor of publications which cite UK institution data at 681 publications, followed by the UK at 336 publications. Collections data enable research that would otherwise be impossible, for instance owing to the scale of information required (Case study 1).

3 https://www.gbif.org/news/5WZThcL928vmPnSvrGhZfE/report-reveals-return-on-investments-in-gbif

## Case study 1 - UK collections data, innovation, and climate change[4]

Scientists with the British Butterfly collection © Trustees of the Natural History Museum, London

Phillip Fenberg, Researcher at the University of Southampton and Science Associate at the Natural History Museum, uses collections data, combined with occurrence records from monitoring and other key datasets such as temperature, to ask questions such as how organisms respond to climate change. Natural science collections enable these questions to be studied over periods of many decades.

While Phillip's original PhD research involved him visiting museum collections to gather specimen data in person (e.g., body size measurements, occurrence records), digital collections data have transformed the efficiency and scope of what is possible. The combination of digital collections images and new computer vision techniques for analysis is incredibly powerful, allowing for previous hypotheses to be tested at scale. For example, Phillip and his team used an NHM dataset of over 180,000 UK and Irish butterfly specimens to show how the adult body size of butterflies responds to warmer temperatures.

Use of an innovative computer vision pipeline - 'Mothra' – showed that it was possible to accurately detect specimens in images, set the scale, measure wing features such as forewing length, and identify the sex. Not only that but like for like comparison of forewing length measurements showed that Mothra could complete work in a week (or less, if more than ten analyses had been run in parallel on a computer cluster) that would take a human some 3,000 hours, or around two years (assuming eight hours a day with no breaks, and only one measurement (forewing length) per specimen).

Phillip is looking forward to the expansion of digital collections image data, particularly the possibilities that will come with increased linkage between genetic and image datasets; the greater integration of artificial intelligence into taxonomic work; and useful metadata such as information on what proportion of any particular collection set has been digitised.

---

4 Interview with Phillip Fenberg 9th January 2023. And see: Rebecca J. Wilson, Alexandre Fioravante de Siqueira, Stephen J. Brooks, Benjamin W. Price, Lea M. Simon, Stéfan J. van der Walt, Phillip B. Fenberg, Applying computer vision to digitised natural history collections for climate change research: Temperature-size responses in British butterflies, 2022, https://doi.org/10.1111/2041-210X.13844

Previous work with Frontier Economics examined the benefits of digitising UK natural science collections in five areas: conservation; invasive species; agricultural research and development; medicines. Looking at current use suggests that these estimates are conservative, confirming how much research is already taking place in these key areas using UK collections data, with over 600 publications on conservation, over 250 on invasive species, and over 100 relating to agriculture (where previous economic benefits estimated at £20-70 million focus on a single use case in relation to wild relatives of food crops). The current study also does not take into account the impact of the UK's important geo-science collections, which are not represented on GBIF, and which inform key areas such as securing sustainable UK supply chains of Lithium[5]. Overall, it is likely that the economic and research efficiency benefits of UK collections data are considerably higher than has yet been demonstrated.

5 https://www.nhm.ac.uk/our-science/our-work/sustainability/li4uk.html

# Case study 2 – collections data and conservation[6]


Herbarium sheet of Ophiocordyceps sinensis © The Board of Trustees of the Royal Botanic Gardens, Kew

Conservation scientist James Westrip is a ‘superuser’ - an author of some 117 papers citing UK collections data, owing to his work with the International Union for Conservation of Nature (IUCN), assessing species for the ‘Red List’ on species conservation status.

Red List assessment demands good data about species distribution over time – many species are data deficient and cannot be assessed, meaning that risks of biodiversity loss are greater than reported, and key conservation actions may be missed.

Since 2019, GBIF data have been transforming how James (and his colleagues) do their work. Geographical data is the most critical for them – ideally in the form of a fully geo-referenced latitude and longitude for specimen collection, but descriptions from labels can be sufficient. This enables species distribution and prevalence to be examined over time, based on different collecting events. Habitat data can also be helpful – one of the benefits of collections data is their coverage of rarer species that are not often observed by humans otherwise.

These data are helping to make the Red List more comprehensive and in particular more representative of species diversity, covering for example more insects, plants and fungi as well as vertebrates which were traditionally well-represented.

And they make the work much more efficient – combining digital specimen data with mapping tools reduces the time taken for many species assessments from weeks to just a day or two. While data quality isn’t always perfect, James has processes to identify and remove outliers. Digital data have reduced the checks needed with collections staff. This work directly informs policy decisions, so the more data are available, and the more species covered, the more impactful it will be.

6 Interview with James Westrip, 6th January 2023

# Unlocking the potential

The demand for, and potential of, UK natural science collections data are very clear. We can see truly remarkable usage, research impact, and benefits for research efficiency, the economy and society even from less than 6% of relevant collections' data.

It is estimated that $44 trillion of economic value generation (or over 50% of the world's GDP) is moderately or highly dependent on nature, with biodiversity loss and ecosystem collapse among the key challenges that the planet faces[7]. Understanding what is in collections now, in the UK and globally, is also key to understanding what is needed as we collect for the future, to underpin policy and investment decisions in future centuries[8].

As the custodians of collections from around the globe, digitisation of natural science collections held in the UK also supports the involvement of communities of origin, and the enrichment of collections through the knowledge and experience of these communities and of experts from the global network. The broader significance of these collections for education, the arts and humanities, and of course leisure and wellbeing can also only be enhanced by the availability of digital collections data for discovery and access.

The UK has set itself the ambition to be a science and technology superpower[9], and natural science collections present an opportunity for the UK to be at the forefront – but while the UK are thought leaders in collections digitisation e.g. the development of data standards and digitisation workflows, we are falling behind[10] in the investment needed to unlock these incredible assets and the value that they can generate, both directly for the UK, and to underpin a future in which both people and planet thrive.

In order to address these challenges and unlock the potential of collections data, UK Natural Science collections are joining forces through the Distributed System of Scientific Collections UK (DiSSCo UK)[11] to set the vision and make the business case for investment in a distributed research infrastructure, working with the Arts & Humanities Research Council (AHRC, who are responsible for heritage collections as research infrastructure) and wider UK Research and Investment (UKRI). This will take the form of digitisation of critical mass of collections as FAIR data, made available through a well-supported backbone of technology infrastructure, with continued development and application of innovative approaches such as greater use of Artificial Intelligence. Investment of the order of £155m is expected to unlock at least a seven- to ten- fold economic return on investment, as well as efficiency savings for UK and global researchers, and the potential for research innovation and studies that have not previously been possible – underpinning a future in which both people and planet thrive.

---

7 Nature Risk Rising: Why the Crisis Engulfing Nature Matters for Business and the Economy, World Economic Forum (2020) https://www.weforum.org/reports/nature-risk-rising-why-the-crisis-engulfing-nature-matters-for-business-and-the-economy/

8 Kirk R. Johnson, Ian F. P. Owens and the Global Collection Group 2023, A global approach for natural history museum collections DOI: 10.1126/science.adf6434

9 https://www.gov.uk/government/publications/the-uk-as-a-science-and-technology-superpower
10 See, for example, recommendation 11 of the 2023 Nurse Review of the RDI landscape https://www.gov.uk/government/publications/research-development-and-innovation-organisational-landscape-an-independent-review
11https://www.dissco-uk.org/