dissco-uk.org

Harnessing the power of the UK's natural science collections

Vince Smith, NHM London
v.smith@nhm.ac.uk, @vsmithuk

NBN Conference 2022, London, UK
10:30-10:45, 9 Nov. 2022

NBN Trust
Making data work for nature

UK collections

Big, diverse, and with global coverage
More than vouchered observations – many uses
Complementary to observation data, but mostly physical

UK collections community (2019)






- Many very significant UK collections
- Limited coordination among UK collections
  - Cf Excellent network of UK observers (e.g., NBN)
- Very diverse governance, strategies & scale
- Limited awareness of European initiatives
- Shared ambition to do more together & a broad appreciation of ‘digital’
- Strong scientific, environmental & societal case for cooperation
- Encouragement from funders (AHRC)

Towards a national (UK) plan

- Initial meetings with larger institutions to develop a common agenda
  - Digitisation
  - Data use (but diverse priorities)
  - Digital infrastructure
  - (Other issues parked)
- Recognition of the need for ‘neutral’ coordination
- Alignment of ambitions with European DiSSCo programme
- Prioritising an integrated business case & data on UK collections
- Initial scoping study (c. £200k) by AHRC
- Annual coordination funding c. £150k by AHRC

The UK Distributed System of Scientific Collections

- A partnership of institutions, working together to harness the digital potential of their collections
- A business case for digitising collections
- A plan (blueprint) on how to make this happen
- Evidence on the size, breadth and use of these collections and associated expertise
- Inclusive in every dimension (e.g. size, geography, collection type)
- Aligned with European DiSSCo activities

EMPOWER THE UK NETWORK OF COLLECTIONS THROUGH DIGITISATION

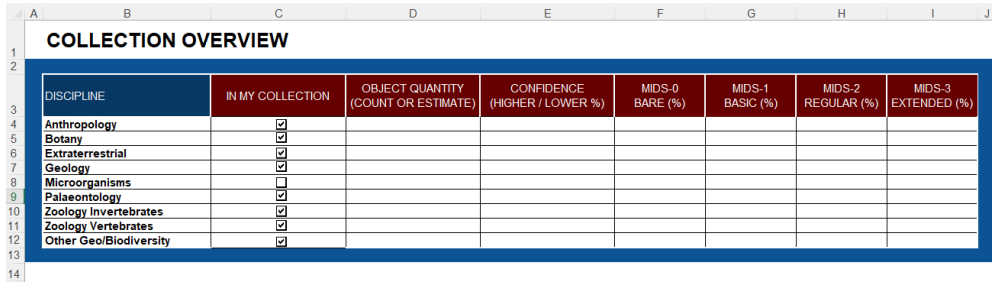
ENHANCE UK BIODIVERSITY AND HUMANITIES INFORMATION INFRASTRUCTURE

IMPROVE DATA QUALITY

DELIVER RELEVANT DATA

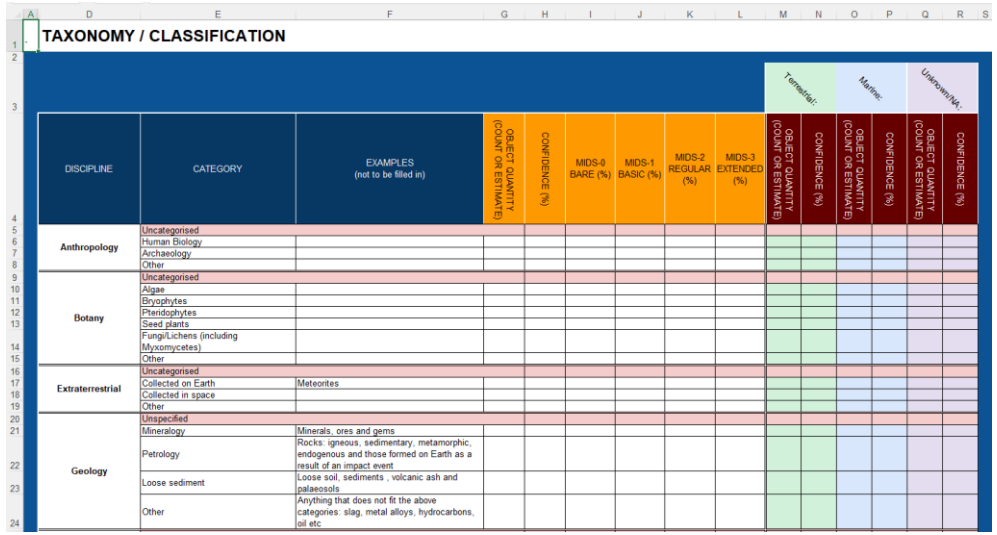
Two UK scoping surveys (Nov 21 - Jan 22)

Collections Survey



COLLECTION OVERVIEW

DISCIPLINE	IN MY COLLECTION	OBJECT QUANTITY (COUNT OR ESTIMATE)	CONFIDENCE (HIGHER / LOWER %)	MIDS-0 BARE (%)	MIDS-1 BASIC (%)	MIDS-2 REGULAR (%)	MIDS-3 EXTENDED (%)
Anthropology	☑						
Botany	☑						
Extraterrestrial	☑						
Geology	☑						
Microorganisms	☐						
Palaeontology	☑						
Zoology Invertebrates	☑						
Zoology Vertebrates	☑						
Other Geo/Biodiversity	☑						

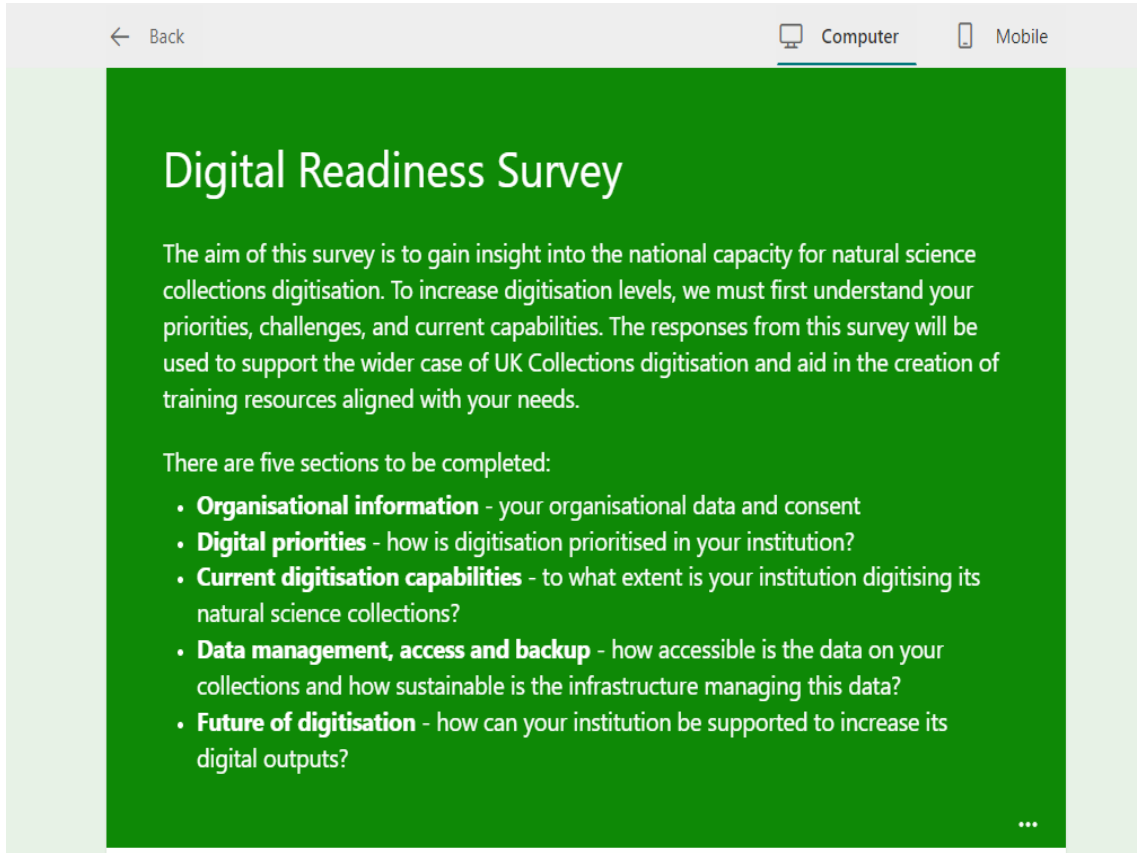


TAXONOMY / CLASSIFICATION

DISCIPLINE	CATEGORY	EXAMPLES (not to be filled in)	OBJECT QUANTITY (COUNT OR ESTIMATE)	CONFIDENCE (%)	MIDS-0 BARE (%)	MIDS-1 BASIC (%)	MIDS-2 REGULAR (%)	MIDS-3 EXTENDED (%)	Terrestrial: OBJECT QUANTITY (COUNT OR ESTIMATE)	Terrestrial: CONFIDENCE (%)	Marine: OBJECT QUANTITY (COUNT OR ESTIMATE)	Marine: CONFIDENCE (%)	Unknown/NA: OBJECT QUANTITY (COUNT OR ESTIMATE)	Unknown/NA: CONFIDENCE (%)
Anthropology	Uncategorised													
	Human Biology													
	Archaeology													
	Other													
Botany	Uncategorised													
	Algae													
	Bryophytes													
	Pteridophytes													
	Seed plants													
	Fungi/Lichens (including Myxomycetes)													
	Other													
Extraterrestrial	Uncategorised													
	Collected on Earth	Meteorites												
	Collected in space													
	Other													
Geology	Unspecified													
	Mineralogy	Minerals, ores and gems												
	Petrology	Rocks: igneous, sedimentary, metamorphic, endogenous and those formed on Earth as a result of an impact event												
	Loose sediment	Loose soil, sediments , volcanic ash and palaeosols												
	Other	Anything that does not fit the above categories: slag, metal alloys, hydrocarbons, oil etc												

What do UK collections hold and to what extent are they already digitised?

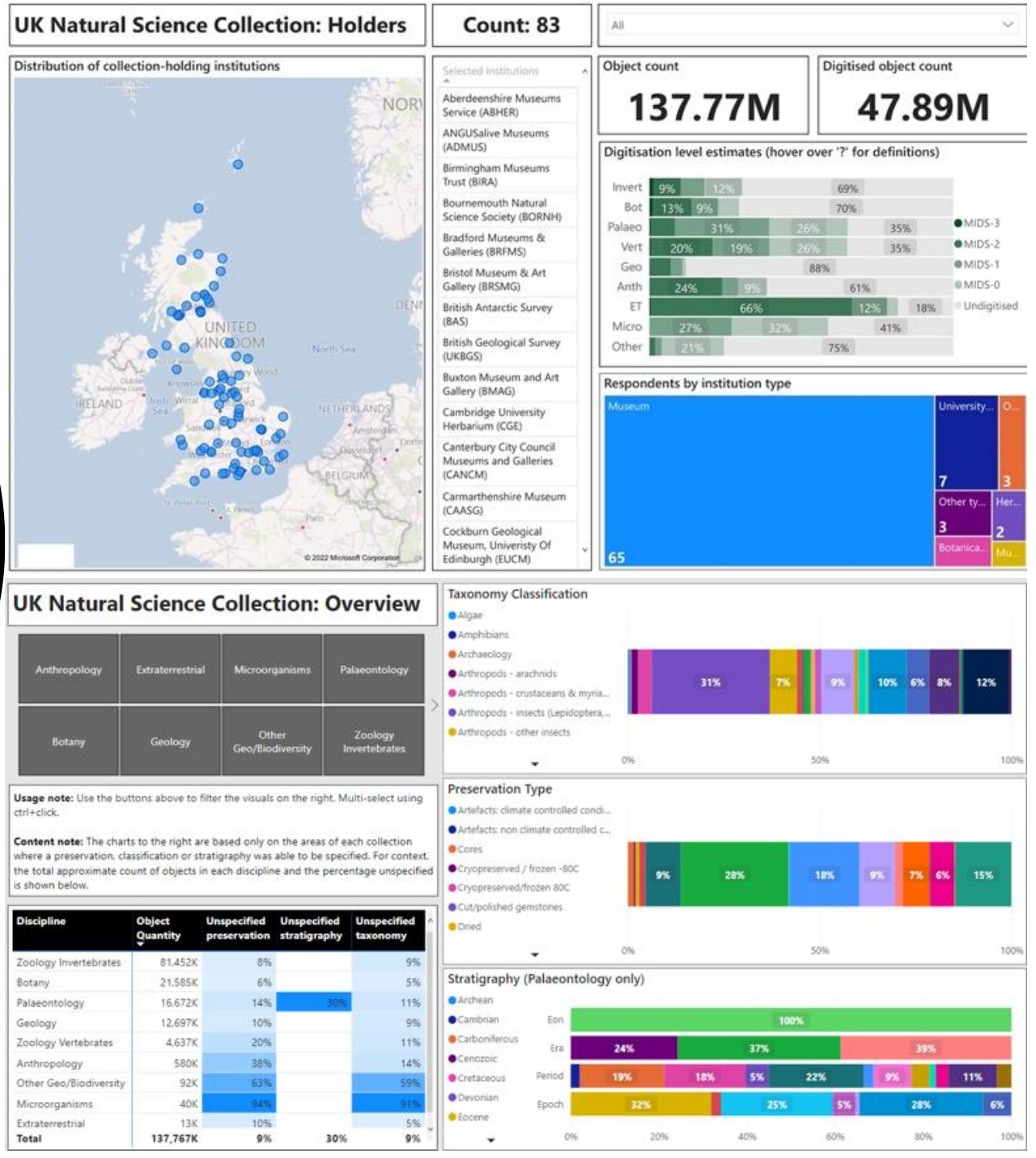
Digital Readiness Survey



What are the current digital capabilities of UK natural science institutions?

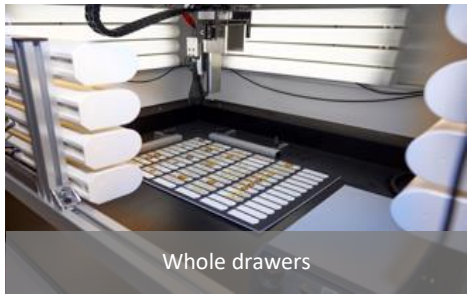
UK Collections

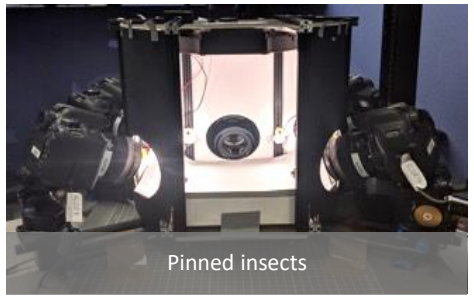
- Interactive dashboard (https://bit.ly/dissco-uk)
- 83 respondents
- Over 137 million specimens
- 59% of these are invertebrates
- 23% specimens have a digital record, but just 2% research ready



UK Digital Readiness

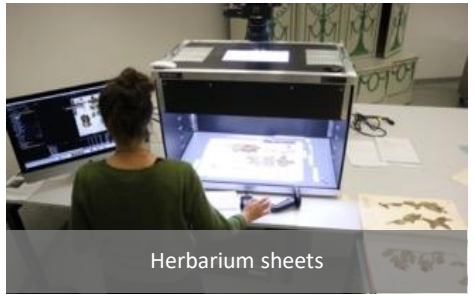
- 90 respondents
- 57% are digitising their natural science collections
- 78% need more funding to prioritise natural science digitisation
- 13% have a full-time digitiser
- 79% don’t systematically share their data


Whole drawers


Pinned insects


Macro fossils


Insect types


Herbarium sheets


Drawers & sheets

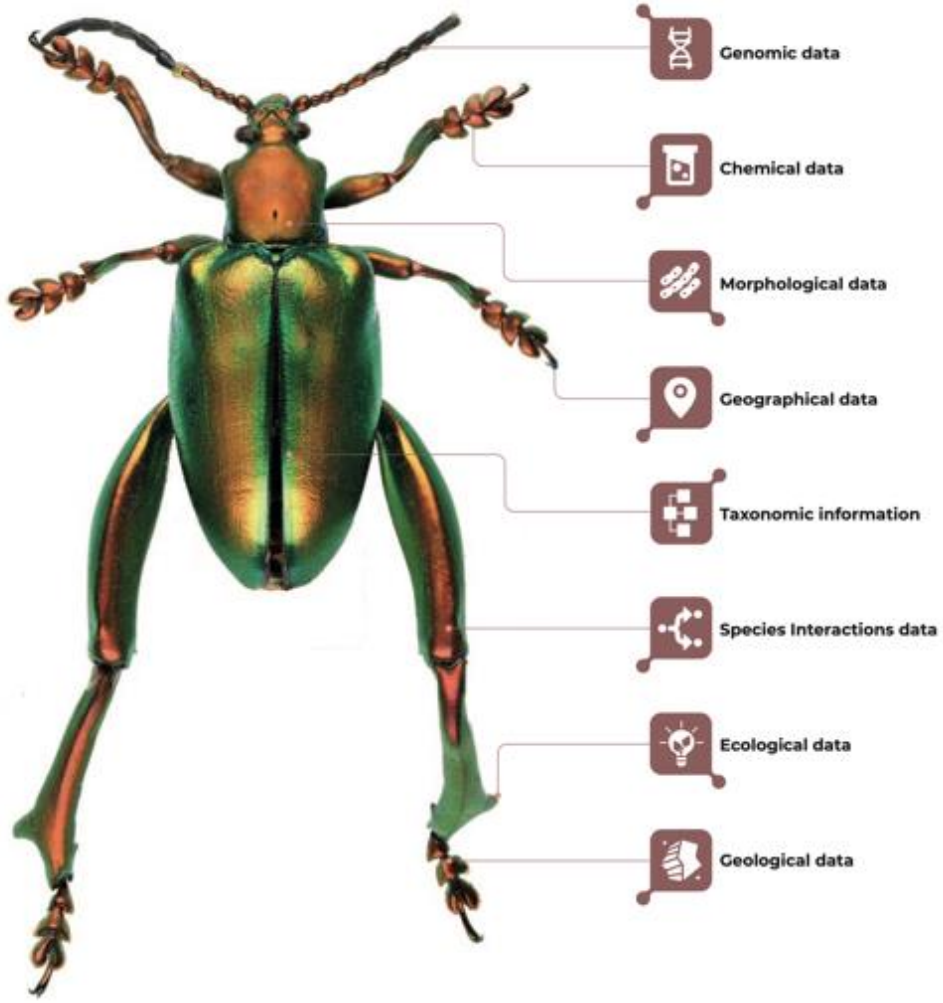

Bound volumes


Surface scans


Multi-light

DiSSCo UK Outputs

- Blueprint for national digitisation programme
- DiSSCo UK website
- Promotional film
- DiSSCo UK Branding guidelines
- DiSSCo Training website additions
- Digitisation equipment for hub pilots
- Global Registry of Scientific Collections (51 institutional records added / updated)
- [separately by NHM] Economic benefits of digitised collections



UK Blueprint

Towards a business case for digitising UK Natural Science Collections

- Part plan, part promotion
- Meeting the needs of a diverse range of stakeholders & institutions
- Emphasising science & infrastructure needs (reflecting funding)
- Inclusive of all UK natural science collections (life & earth, large & small)
- Recognises that our plans are in development
- 28 pages with 5 key sections
- 4 case studies from partners

Benefits of UK NSCs

Why the UK collections are special & what value does digitisation unlock

- Scientific benefits
- Policy impacts
- Educational benefits
- Cultural / societal benefits
- Economic benefits

ADVANCING SCIENCE, INNOVATION AND UK POLICY

Lifelong learning

Managing natural science collections data requires skills and competencies that align with next-generation science standards for all age groups. This applies in both formal educational settings, such as science curricula, as well as informal learning opportunities. The digital data and specimens central to DiSSCo UK can be seamlessly incorporated into existing courses that include topics on evolution, biodiversity, systematics, conservation, climate change and ecology. Specimen-based data makes science accessible through the specimen itself, which is a tangible, place-based, engaging object, as well as through aggregated specimen data that is verifiable, relevant, and a logical gateway to data literacy.

DiSSCo UK cannot reach its full potential without the strong involvement of the citizen science community. The UK has a long history of amateur natural scientists, and there are many UK based citizen science projects already structured around monitoring biodiversity, such as the biological recording schemes and digital platforms like iSpot, iRecord and iNaturalist.

Related digital projects will play a key role in involving the public with DiSSCo UK through supporting involvement with collections-based science and databases. Contributions include transcription, taking measurements and annotating features such as flowering stages. UK collections, both small and large, are in a unique position to attract volunteers to help with the digitisation process, providing valuable training that is inclusive and engages participants from a wide range of ages, interests and backgrounds.

Growing UK collection capabilities

UK natural science collections in recent years have undergone some rationalisation, with some collections physically merged. Local collections are vulnerable to lack of support due to changes in institutional priorities. Most local and regional institutions lack a specialist natural science curator, or share a peripatetic curator. All involve curators in a range of tasks focused not only on collections management but also on visitor engagement, research, item loans, exhibitions and more. DiSSCo UK is an unparalleled opportunity to reinvest in UK collections and the expertise of the staff that care for them.

With collections becoming digital, specimens held at different sites can be managed online as a single entity via the national data portal, and searched in a unified way. Taxonomic expertise can be better coordinated, improving the resolution and spread of taxonomic skills in the UK. The voluntary sector, with its core of expert amateur naturalists, is an important repository of taxonomic expertise and by opening up UK collections digitally, DiSSCo UK has the potential to inspire a new generation of taxonomists amongst the volunteers who monitor changes in their local fauna and flora.

DiSSCo UK's technological innovation will create a powerful set of tools for UK researchers. While the state of collections in the UK is not susceptible to a "quick technological fix", technology is a driver that can be used to support the huge demand for taxonomic information created by biodiversity loss and global change, while also addressing the expectation amongst scientists and the public for information sources that are both accessible and easy to use.

THE ECONOMIC VALUE OF DIGITISING COLLECTIONS

In 2021, the Natural History Museum, London, collaborated with economic consultants Frontier Economics to explore the economic and societal value of digitising natural science collections. They concluded that digitisation has the potential to see a tenfold return on investment, creating benefits in excess of £2 billion over 30 years. The report examined the impact of collection data in five sectors: biodiversity conservation, invasive species, medicine discovery, agricultural research and development, and mineral exploration.

For more information, see https://doi.org/10.3897/rio.7.e78844.

BIODIVERSITY CONSERVATION
£0.7BN TO £1BN
Digitisation enhances taxonomic knowledge which improves detection of threatened species. This enables conservation efforts and maintains balance in the ecosystem.

INVASIVE SPECIES
£0.7BN TO £1.1BN
Digitisation enhances taxonomic knowledge which improves detection of invasive species. Reducing the frequency of losses from threats leads to significant economic benefits.

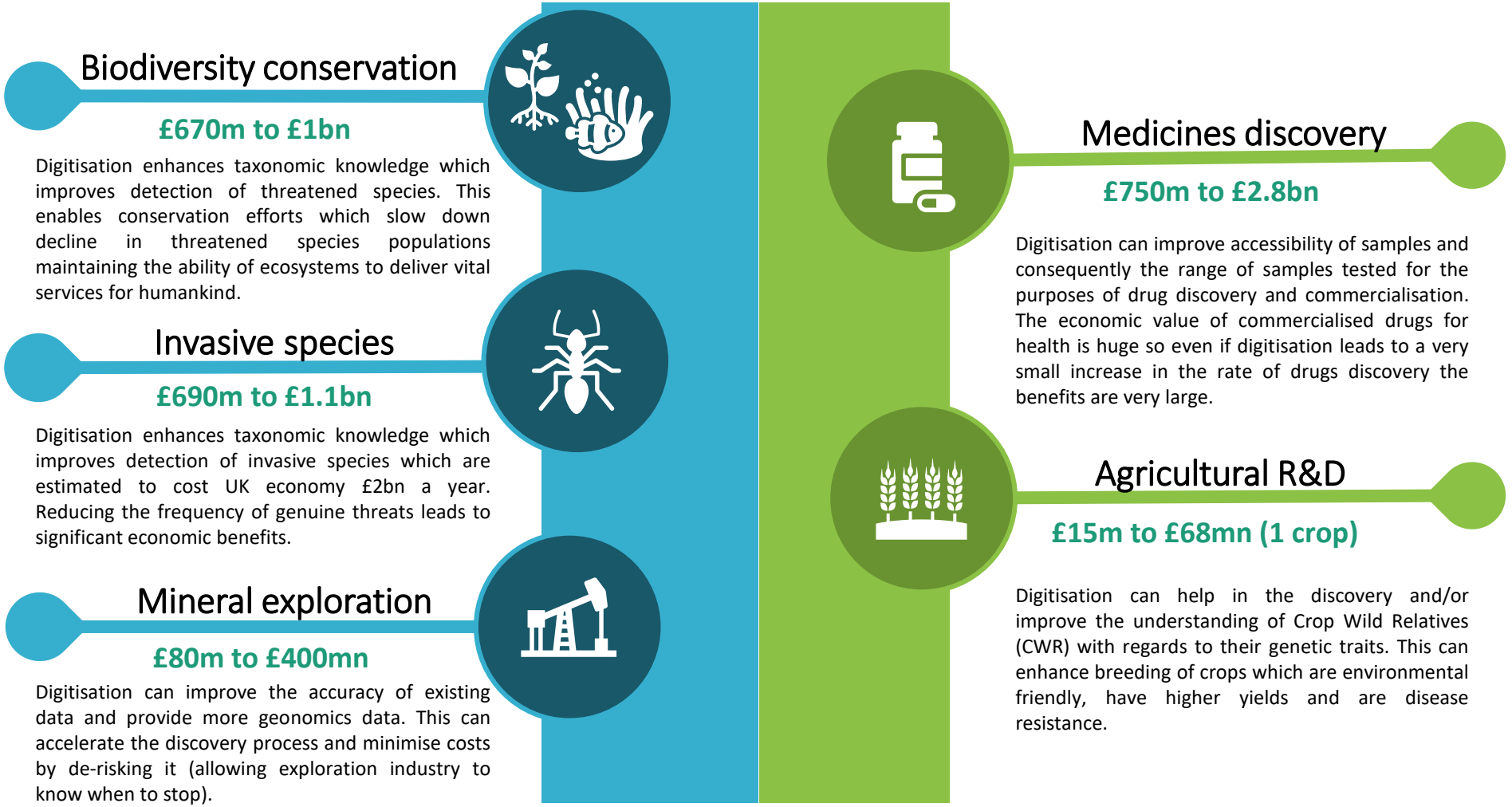
MEDICINES DISCOVERY
£0.8BN TO £2.8BN
Digitisation can increase the availability of samples which can be tested for the purposes of drug discovery.

AGRICULTURAL R&D
£20M TO £70M
Digitisation can help in the discovery and/or improve the understanding of Crop Wild Relatives (CWR) with regards to their genetic traits. This will benefit breeding of better crops for agriculture.

MINERAL EXPLORATION
£0.7BN TO £1BN
Digitisation can improve the accuracy of existing data and provide more geonomics data. This can accelerate the discovery process and minimise costs by de-risking it.

Economic benefits of digitising

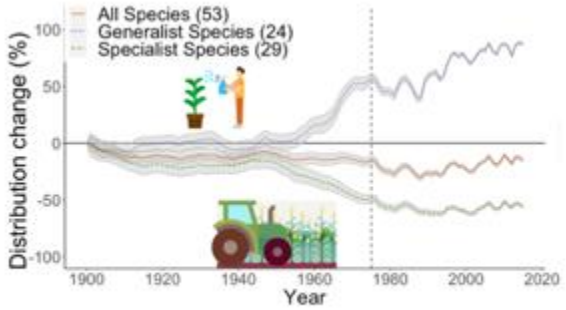
Popov D et al. 2021. RIO, https://doi.org/10.3897/rio.7.e78844



Theory of change methodology, five areas of investigation very conservatively give £2.2bn in benefits

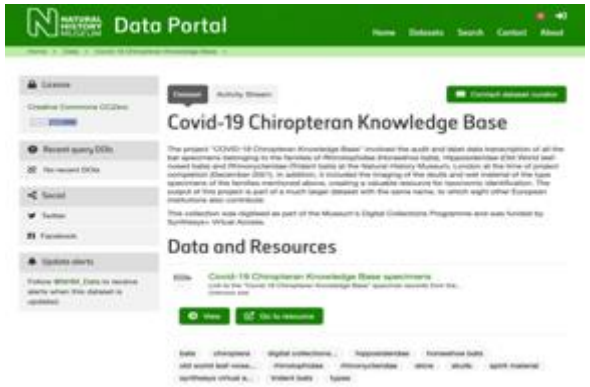
Typical research use cases

Measuring Change

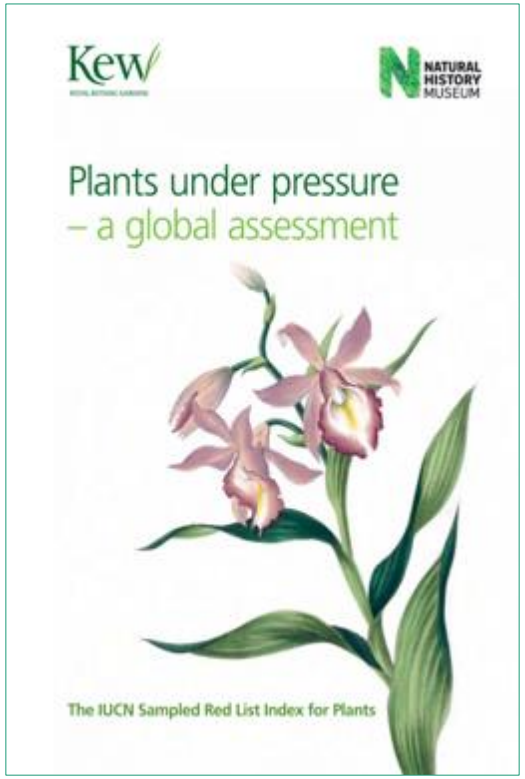


A graph showing rates of butterfly distribution change over the last century

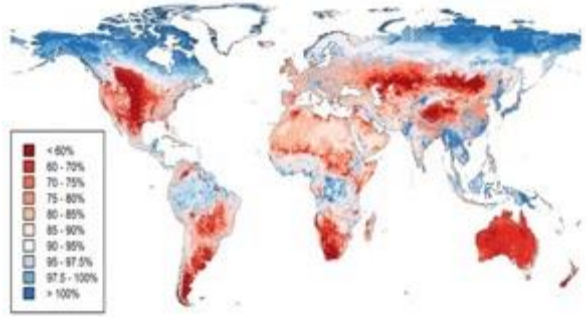
Pandemic Preparedness



Conservation



Biodiversity Indicators



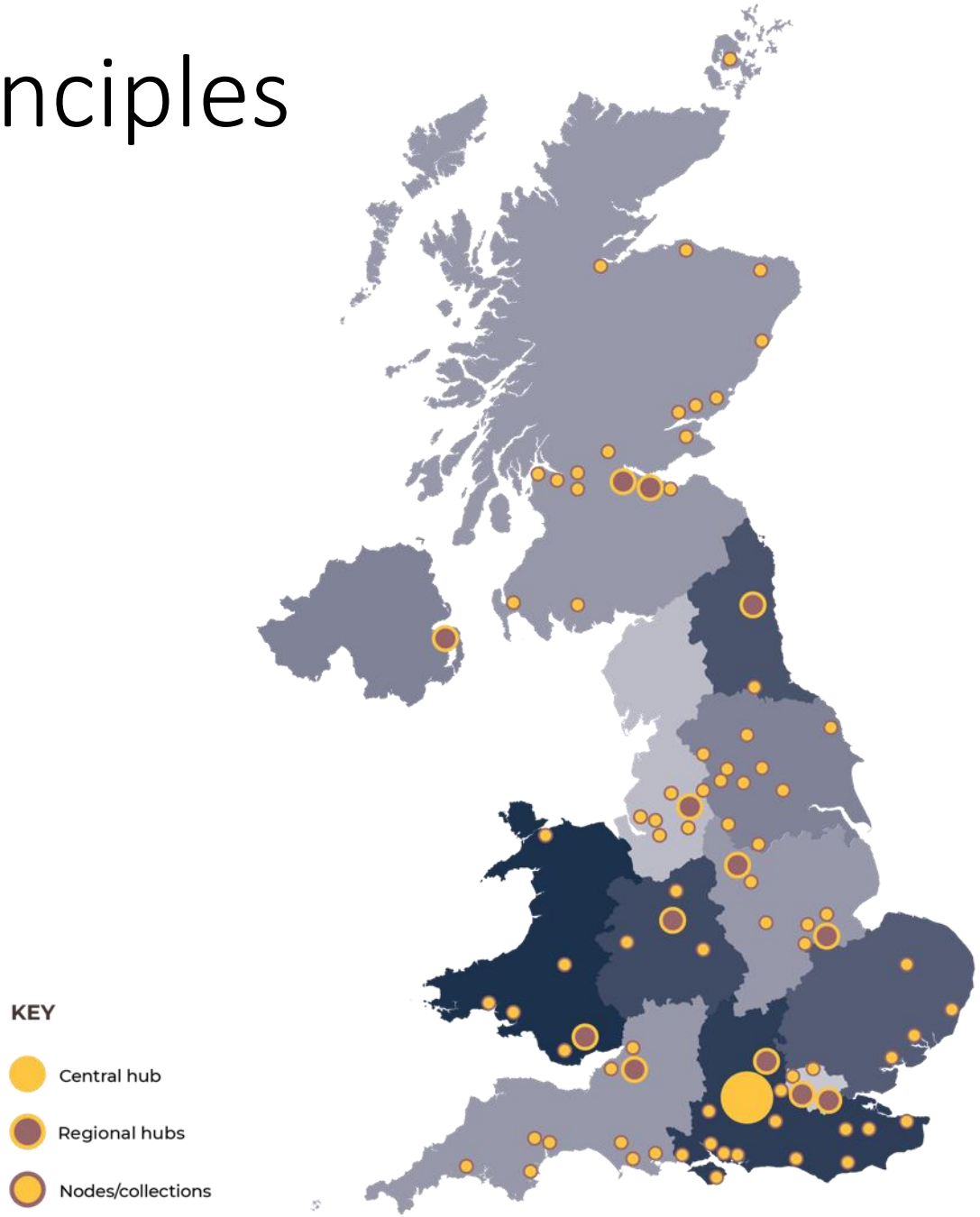
NHM data use metrics: 5m records, 34bn downloads, 600k datasets, 2197 publications (7th Nov. 2022)

DiSSCo UK organisation principles
Hub & spoke model

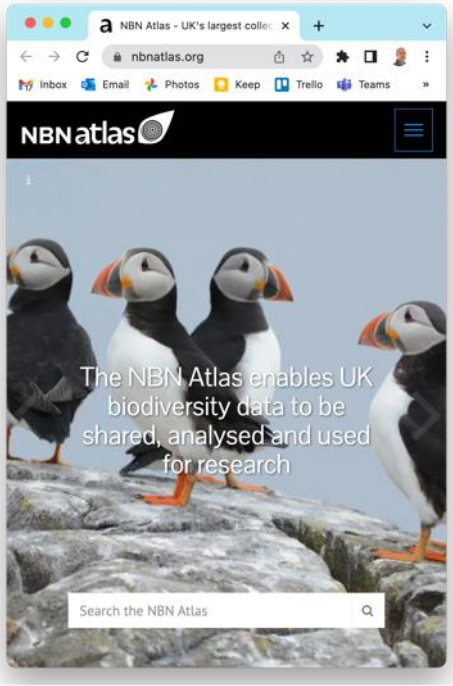
Three tiers:

- Central Hub
  - The coordination point for DiSSCo UK
  - Digitizing their collections & supporting regional hubs
- Regional / thematic hubs
  - Regional/domain community management
  - Digitizing their collections & supporting project partners
  - Emerging role for data partners (CryoArks & NBN)
- Collection nodes
  - Local / intermittent participants

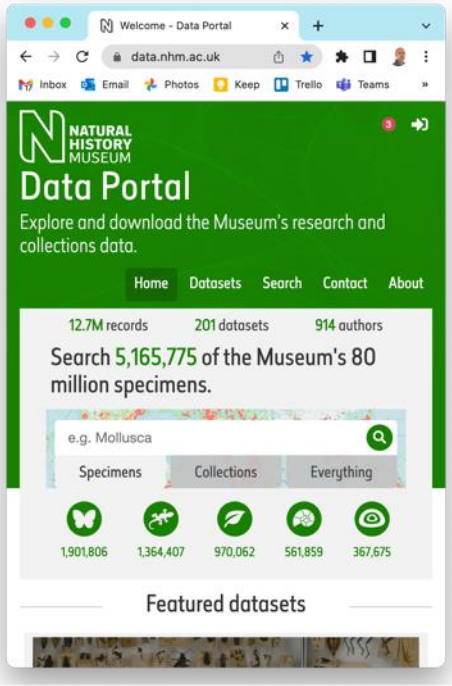
- *Working toward all levels signing the DiSSCo MoU*
- *Sustained programmes of institutional digitization by central and regional/thematic hubs*
- *Project level digitization with local collections*
- *National data aggregation, feeding GBIF & DiSSCo*



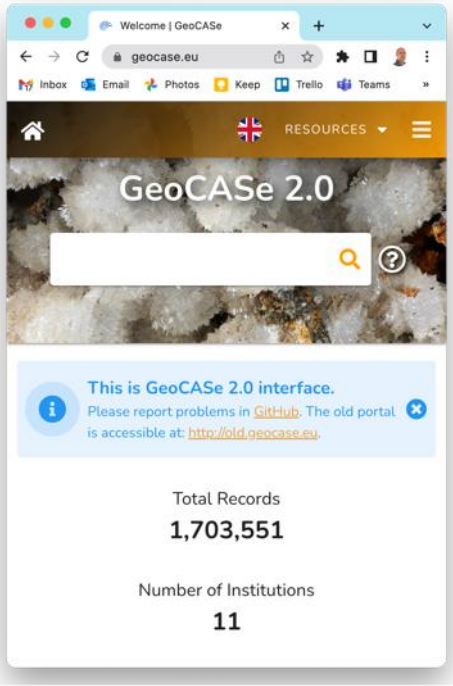
Evaluating national data portal technologies:
one size does not fit all



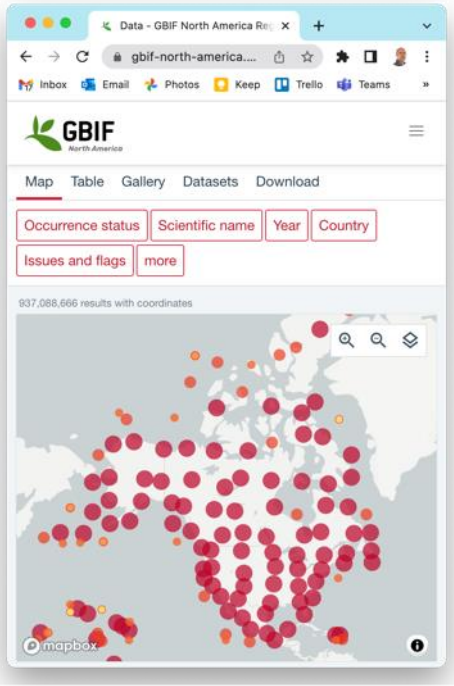
Living Atlas
Used by NBN
National focus
Life science
Rich functionality



CKAN
Used by NHM
Approaching end of life
Life + Earth
Limited functionality

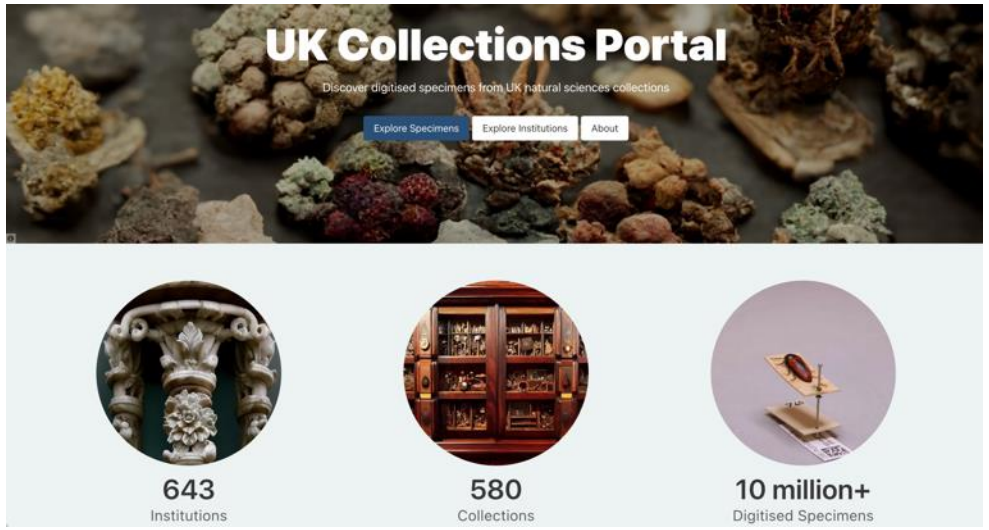


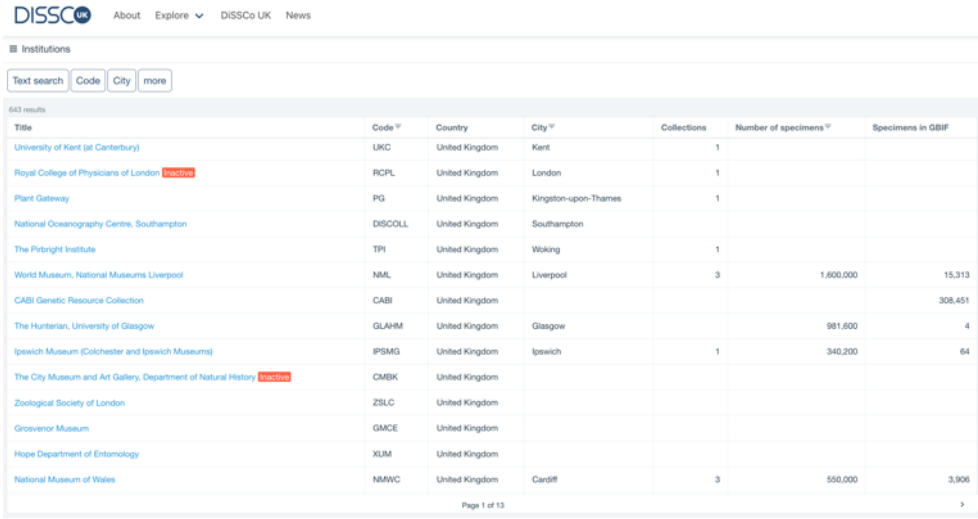
GeoCase
Custom development
High maintenance
Earth only
Limited functionality

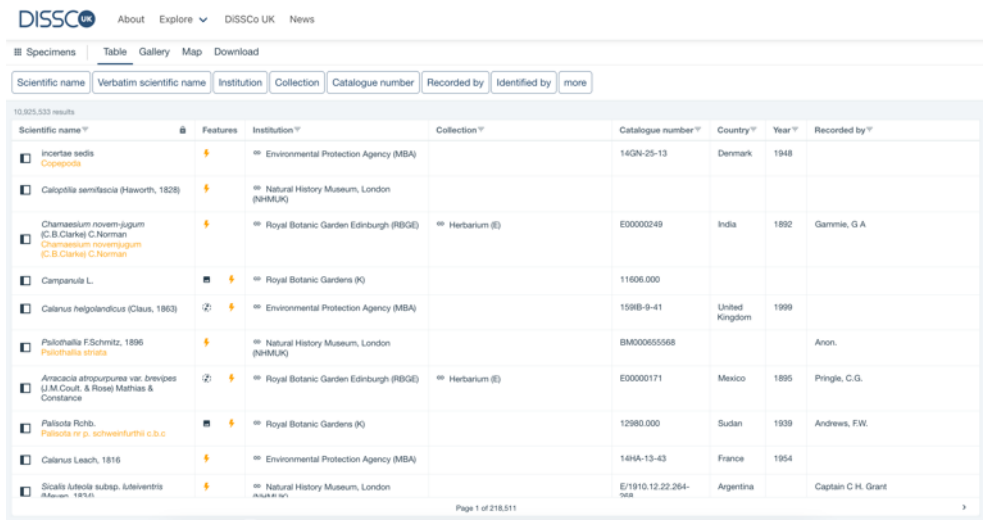


GBIF hosted portal
Rich functionality
Low maintenance
Life + Palaeo only
Integrated with GBIF

Life Science Collections via GBIF Hosted Portal







Application to site in 2 weeks!

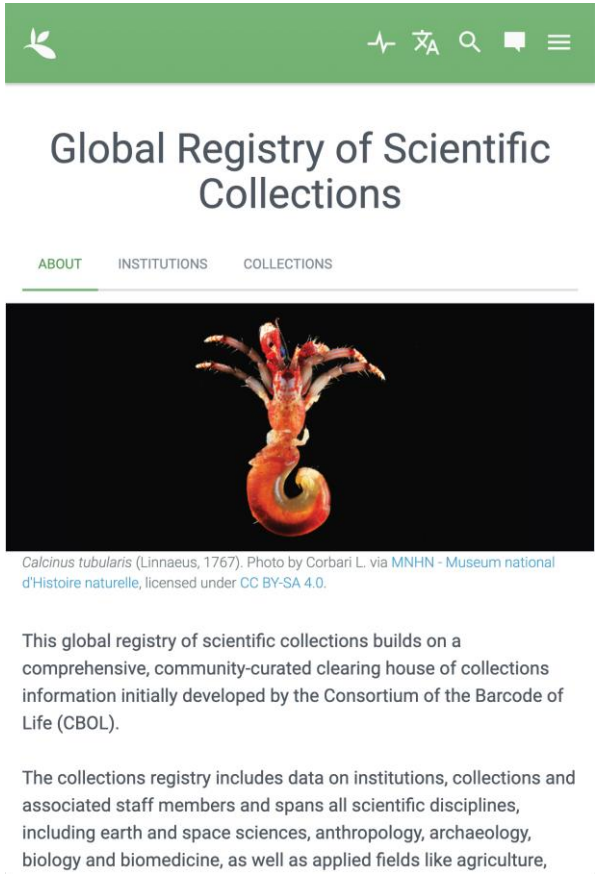
- 643 UK Institutions (many inactive)
- 580 Collections
- 10 million + specimens
- Launching soon

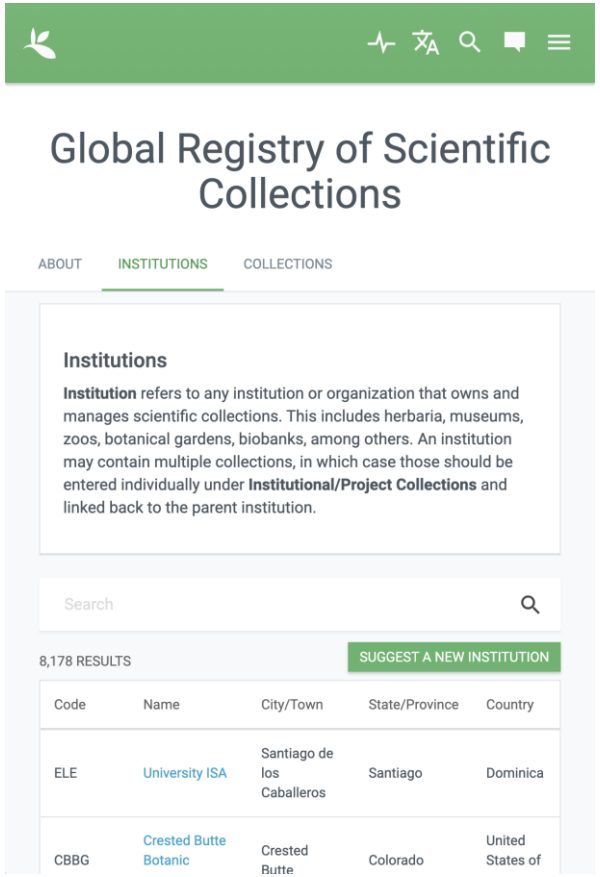
GBIF.org

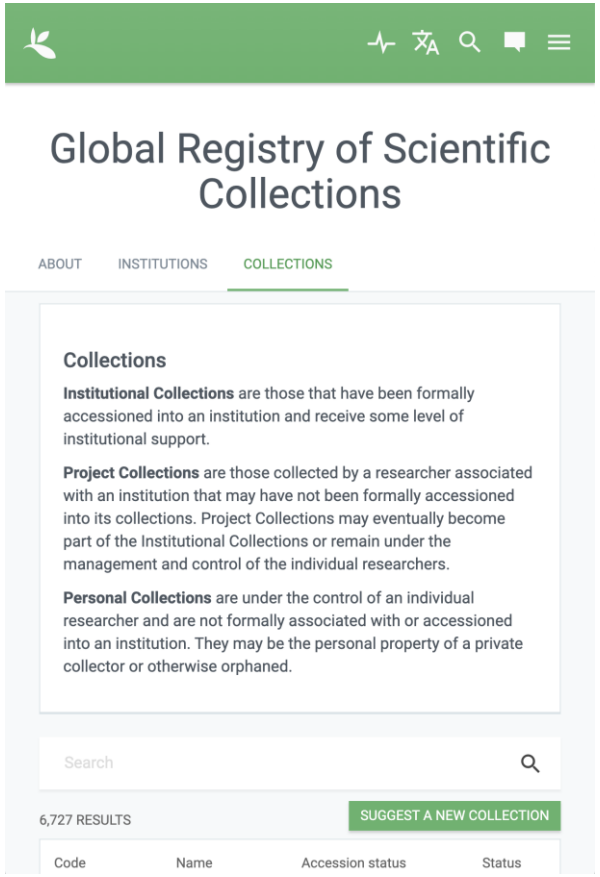
Earth science considerations

- Diverse geological collection types and needs
  - Cores, fossils, rocks, minerals, meteorites, ocean bottom deposits, crystals & gems
- No complete solution available
- GeoCASE Portal is closest
- Poor understanding of all our community needs
- Even less understanding of users outside the collections community
- Working closely with BGS
- Business analyst to report
- Aiming to make recommendations Q1 of ‘23

UK institutions in the Global Registry of Scientific Collections (GRSciColl)

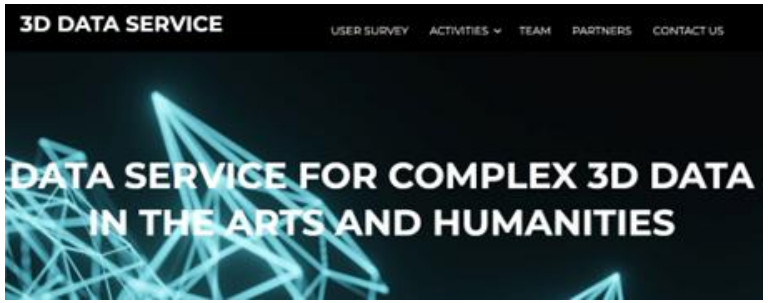


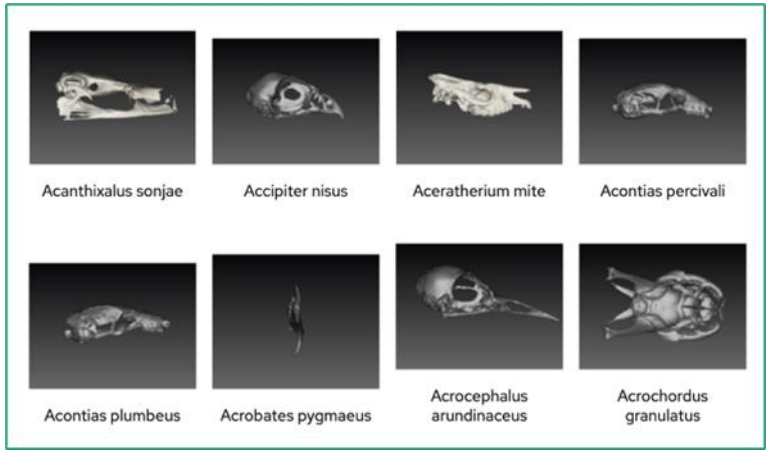




DiSSCo UK Is helping to coordinate improvement of GRSciColl entries for the UK

3D requirements: a UK MorphoSource node?





- Potential for a national resource alongside life & earth data portals
- NHM London has very substantial 3D collections - (Phenome10k)
- NSF/BBSRC joint EoI to develop a Morphosource node
- Potential to become DiSSCo UK node

MorphoSource
3D Data Repository

Three emerging lines of funding…

1. Data repositories & community development (current)

- c. 150k annually (AHRC, iDAH programme)
- DiSSCo UK community activities
- Associated data repository development

2. National digitisation funding (anticipated)

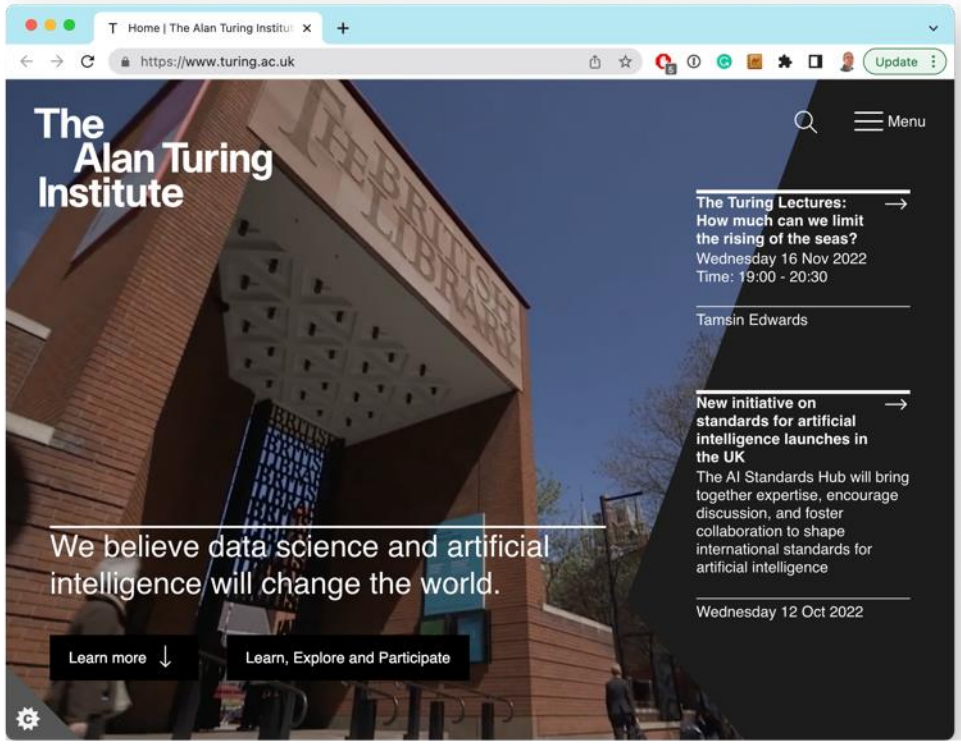
- Starting from 2023 (c. tens of millions)
- Funds content (data) creation, via hubs & nodes, feeding the repositories
- Capacity building and demand led digitization projects at DiSSCo UK nodes

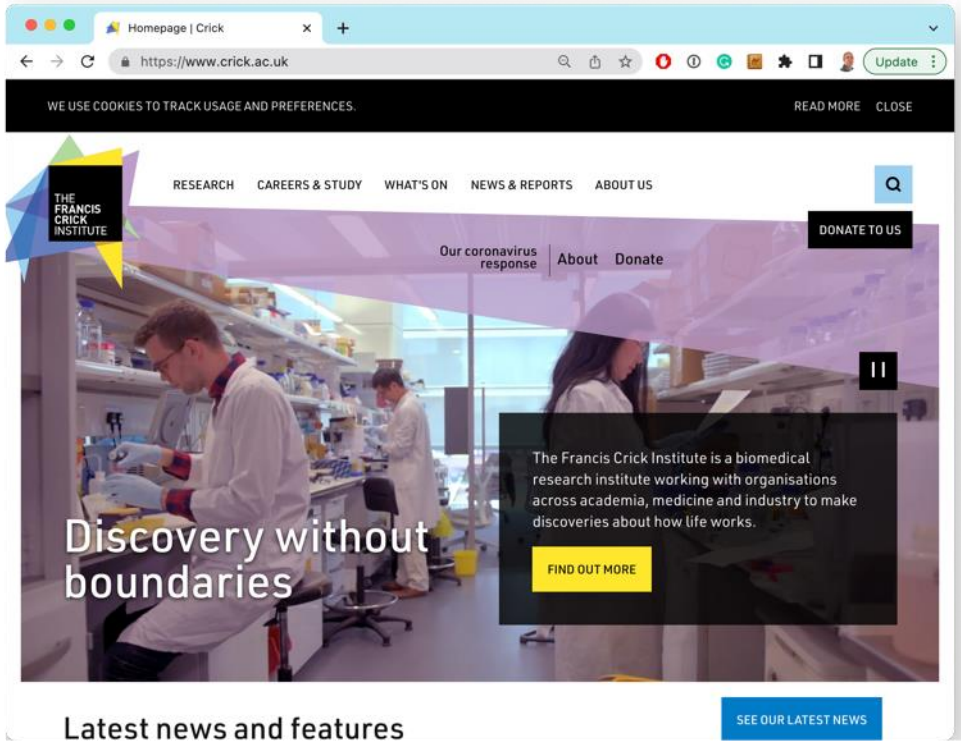
3. New biodiversity institute (proposed)

- In discussion with government, focused on data assimilation & exploitation
- Working across UKRI agencies, integrating big data & AI/ML, genetics
- A possible coordination point for DiSSCo UK?

UKRI Arts and Humanities Research Council

Recent UK research centres…





Alan Turing Institute

Data Science and AI

c. 400 networked researchers across UK

Francis Crick Institute

Biomedical research

c. 1500 staff


DISSC UK
Q&A