Sharing Digital Sequence Information

PRESENTED BY:
Paul Oldham: May 31, 2022

What are we talking about?

- Digital sequence information (DSI) is a place holder term for what has variously been called genetic information, bioinformation, sequence information, natural information, genetic sequence data, nucleotide sequence data or genetic resources *in silico*;
- The definition of digital sequence information is contested (nucleotides, proteins, metabolites etc.) and not terribly constructive (salami slicing);
- DSI is being negotiated under multiple instruments and processes (CBD, Nagoya, UNCLOS, FAO, WHO)
- Countries may adopt increasingly restrictive approaches to sharing digital sequence information

The Central Puzzle

- A lot of work has been taking place to try and understand what is happening and possible options (studies, articles, webinars etc)
- The central puzzle is: **how to keep dsi open access and generate monetary benefits for investments in biodiversity?**
- In my work: two studies for the European Commission*
  - technical aspects of DSI (2020)
  - sharing digital sequence information (with Jasmine Kindness, 2022)

*Disclaimer: The views expressed in the studies are entirely the author's own and do not reflect the official opinion or views of the European Commission

A Skeleton Proposal

A cross instrument approach with 6 main elements:

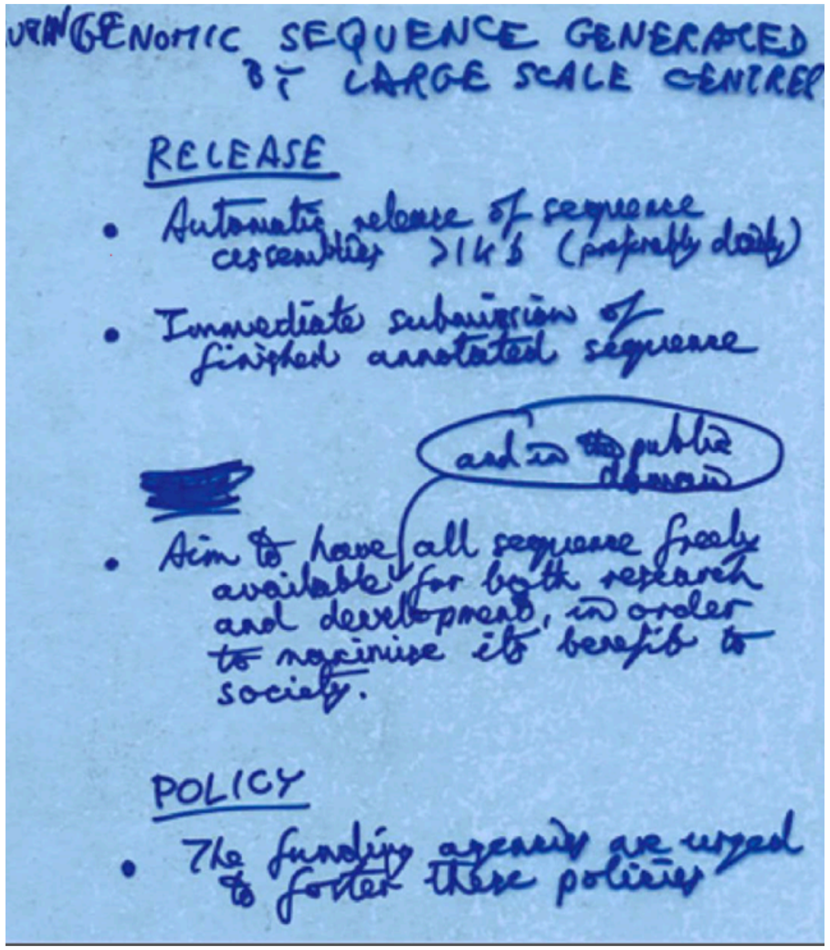
1. Principles
2. Open Licences
3. Revenue Generating Measures (models)
4. A Distributed Global Biodiversity Fund (DGBF)
5. A Shared Roadmap (to operationalise post 2020 GBF)
6. Participants

1. Principles

- Open access and open science (UNESCO 2021 Recommendation on Open Science);
- Enhancing access to and sharing of data (OECD Council 2021 recommendation on Enhancing Access to and Sharing of Data);
- Human rights based approach (IPLCs, UNDRIP etc., Responsible Research & Innovation policies);
- Embed financial benefits for Nature in the policy solution (Dasgupta 2021, The Economics of Biodiversity);
- Operationalise the post-2020 GBF etc…

2. Open Licences

Human Genome Project



GENOMIC SEQUENCE GENERATED BY LARGE SCALE CENTRES

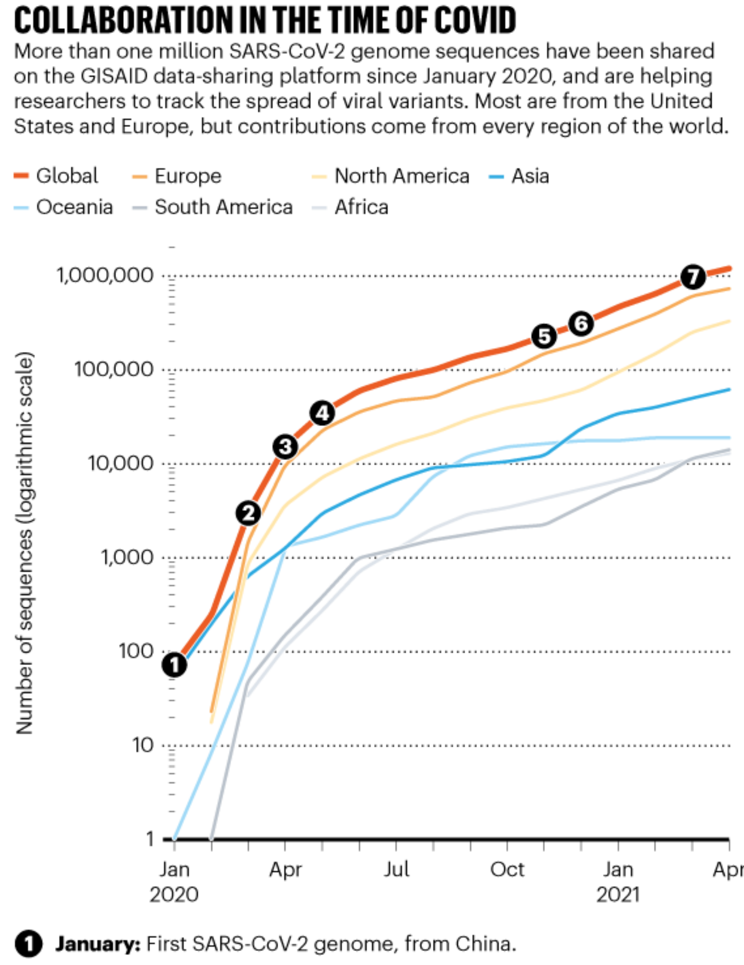
RELEASE

- Automatic release of sequence assemblies >1kb (preferably daily)
- Immediate submission of finished annotated sequence and in the public domain
- Aim to have all sequence freely available for both research and development, in order to maximise its benefit to society.

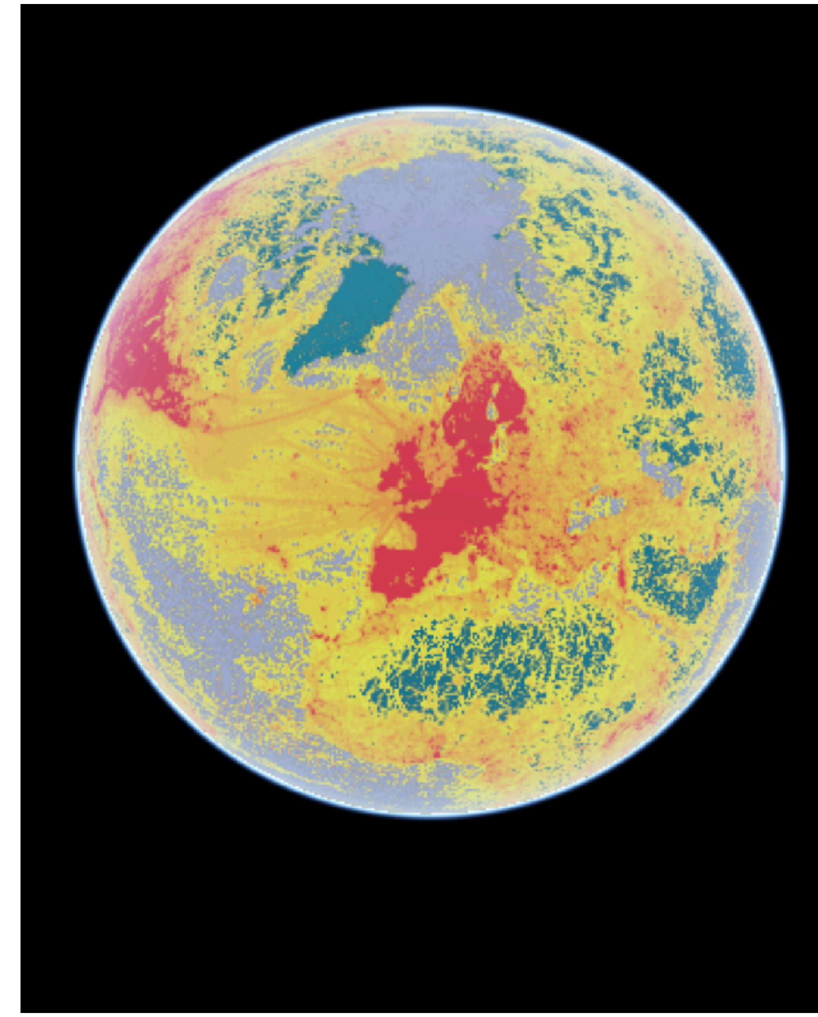
POLICY

- The funding agencies are urged to foster these policies

Influenza & COVID 19



Taxonomic Data



INSDC
Free for All

- 1996 Bermuda Principles focus on release of humane genome sequence into the public domain
- Becomes a *de facto* norm in the 2000s for all genomes (Fort Lauderdale Agreement 2003)
- Users have no obligations or responsibilities to providers

GISAID
Closed Commons

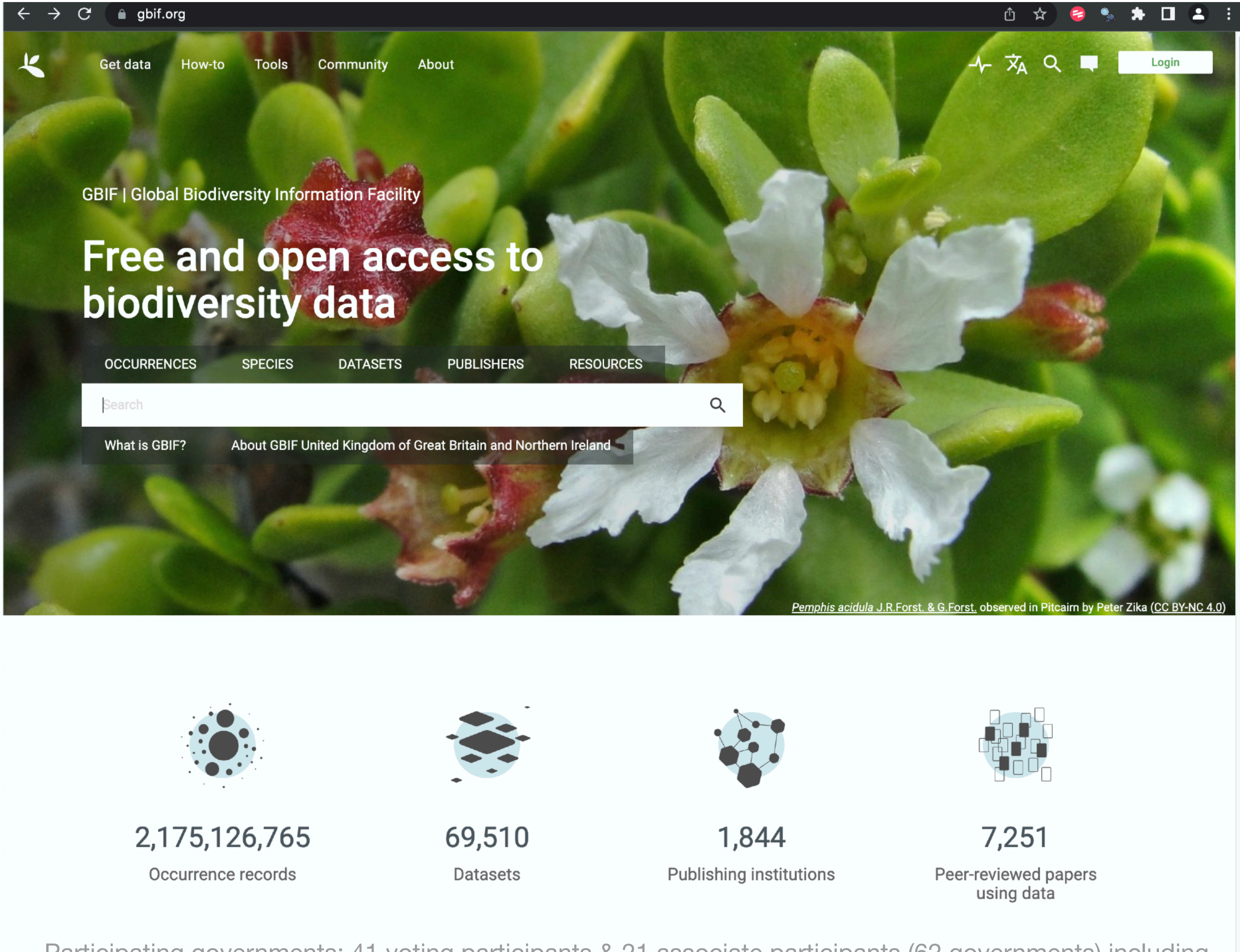
- Response to problems with sharing Bird flu (H1N1) samples
- T&Cs require "best efforts" in collaboration. No sharing of sequence outside the closed commons
- Very popular but unlikely to scale for all dsi globally

GBIF
Open Commons

- Intergovernmental network
- Confronted by complex licence provisions
- 2014 consultation leads to the adoption of three Creative Commons licences
- Applies to 2.1 billion taxonomic records including sequence metadata from INSDC

Three Data Sharing Models

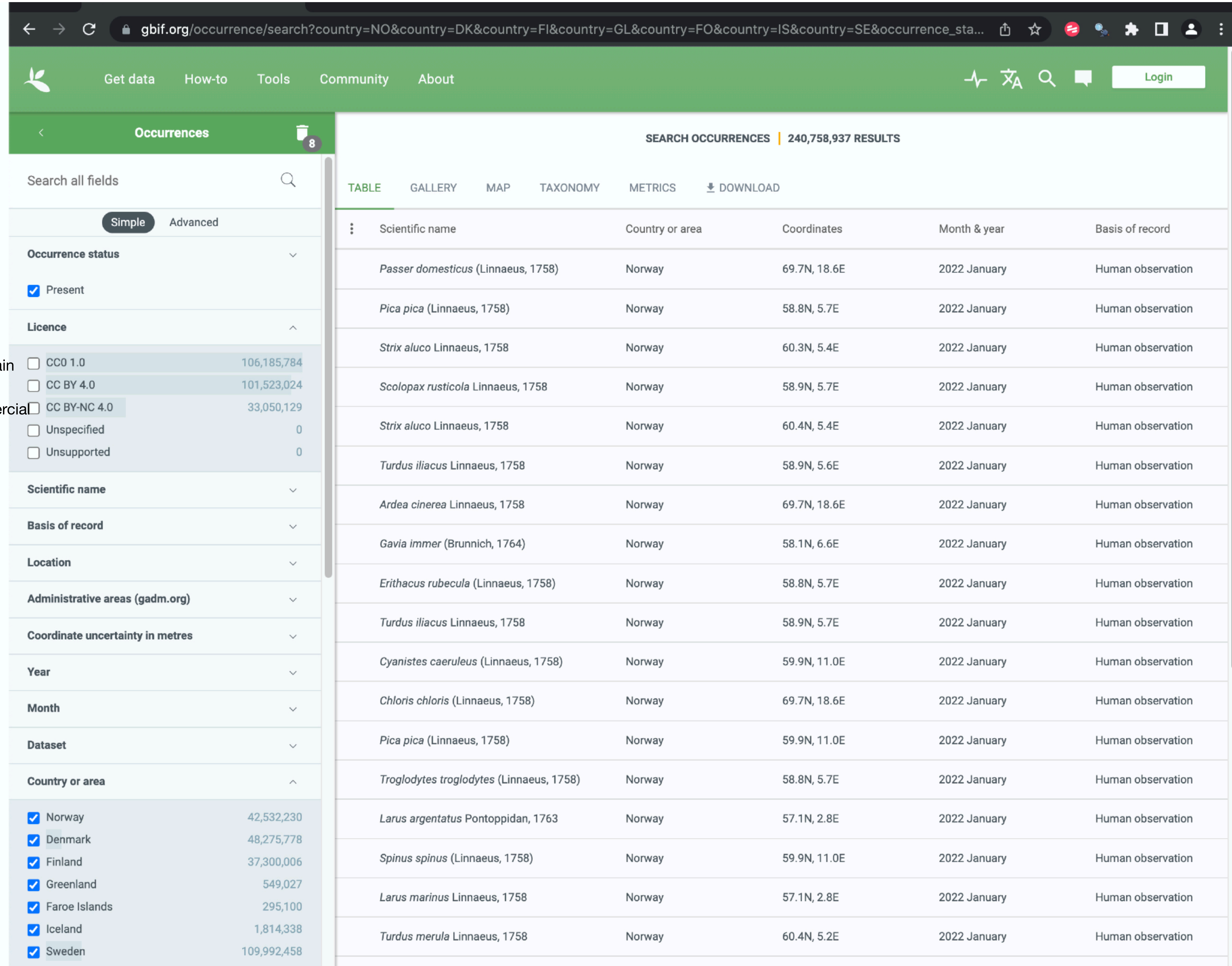
https://www.gbif.org/



Participating governments: 41 voting participants & 21 associate participants (62 governments) including non-Parties to the CBD. INSDC & EMBL-EBI as publishers.

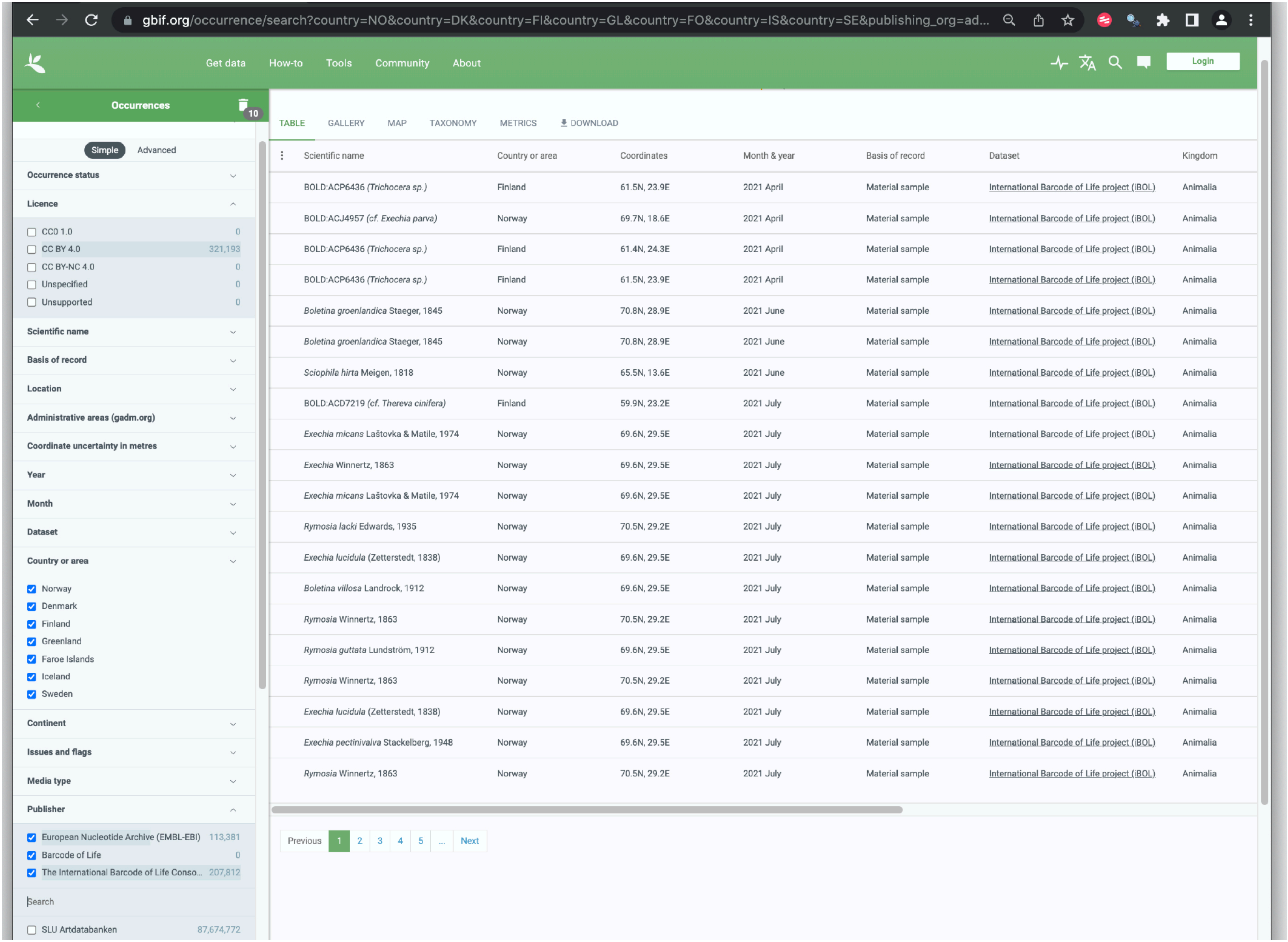
Nordic Council Country Records (240 million records)

- Public domain
- Attribution
- Non-commercial



Visit this data here

321,000 sequence metadata records for Nordic Countries (attribution licence)


gbif.org/occurrence/search?country=NO&country=DK&country=FI&country=GL&country=FO&country=IS&country=SE&publishing_org=ad...

Get data How-to Tools Community About Login

Occurrences 10

Simple Advanced

Occurrence status

Licence

- CC0 1.0 0
- CC BY 4.0 321,193
- CC BY-NC 4.0 0
- Unspecified 0
- Unsupported 0

Scientific name

Basis of record

Location

Administrative areas (gadm.org)

Coordinate uncertainty in metres

Year

Month

Dataset

Country or area

- Norway
- Denmark
- Finland
- Greenland
- Faroe Islands
- Iceland
- Sweden

Continent

Issues and flags

Media type

Publisher

- European Nucleotide Archive (EMBL-EBI) 113,381
- Barcode of Life 0
- The International Barcode of Life Conso... 207,812

Search

- SLU Artdatabanken 87,674,772

TABLE GALLERY MAP TAXONOMY METRICS DOWNLOAD

Scientific name	Country or area	Coordinates	Month & year	Basis of record	Dataset	Kingdom
BOLD:ACP6436 *(Trichocera sp.)*	Finland	61.5N, 23.9E	2021 April	Material sample	International Barcode of Life project (iBOL)	Animalia
BOLD:ACJ4957 *(cf. Exechia parva)*	Norway	69.7N, 18.6E	2021 April	Material sample	International Barcode of Life project (iBOL)	Animalia
BOLD:ACP6436 *(Trichocera sp.)*	Finland	61.4N, 24.3E	2021 April	Material sample	International Barcode of Life project (iBOL)	Animalia
BOLD:ACP6436 *(Trichocera sp.)*	Finland	61.5N, 23.9E	2021 April	Material sample	International Barcode of Life project (iBOL)	Animalia
Boletina groenlandica Staeger, 1845	Norway	70.8N, 28.9E	2021 June	Material sample	International Barcode of Life project (iBOL)	Animalia
Boletina groenlandica Staeger, 1845	Norway	70.8N, 28.9E	2021 June	Material sample	International Barcode of Life project (iBOL)	Animalia
Sciophila hirta Meigen, 1818	Norway	65.5N, 13.6E	2021 June	Material sample	International Barcode of Life project (iBOL)	Animalia
BOLD:ACD7219 *(cf. Thereva cinifera)*	Finland	59.9N, 23.2E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia
Exechia micans Laštovka & Matile, 1974	Norway	69.6N, 29.5E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia
Exechia Winnertz, 1863	Norway	69.6N, 29.5E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia
Exechia micans Laštovka & Matile, 1974	Norway	69.6N, 29.5E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia
Rymosia lacki Edwards, 1935	Norway	70.5N, 29.2E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia
Exechia lucidula (Zetterstedt, 1838)	Norway	69.6N, 29.5E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia
Boletina villosa Landrock, 1912	Norway	69.6N, 29.5E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia
Rymosia Winnertz, 1863	Norway	70.5N, 29.2E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia
Rymosia guttata Lundström, 1912	Norway	69.6N, 29.5E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia
Rymosia Winnertz, 1863	Norway	70.5N, 29.2E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia
Exechia lucidula (Zetterstedt, 1838)	Norway	69.6N, 29.5E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia
Exechia pectinivalva Stackelberg, 1948	Norway	69.6N, 29.5E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia
Rymosia Winnertz, 1863	Norway	70.5N, 29.2E	2021 July	Material sample	International Barcode of Life project (iBOL)	Animalia

Previous 1 2 3 4 5 ... Next

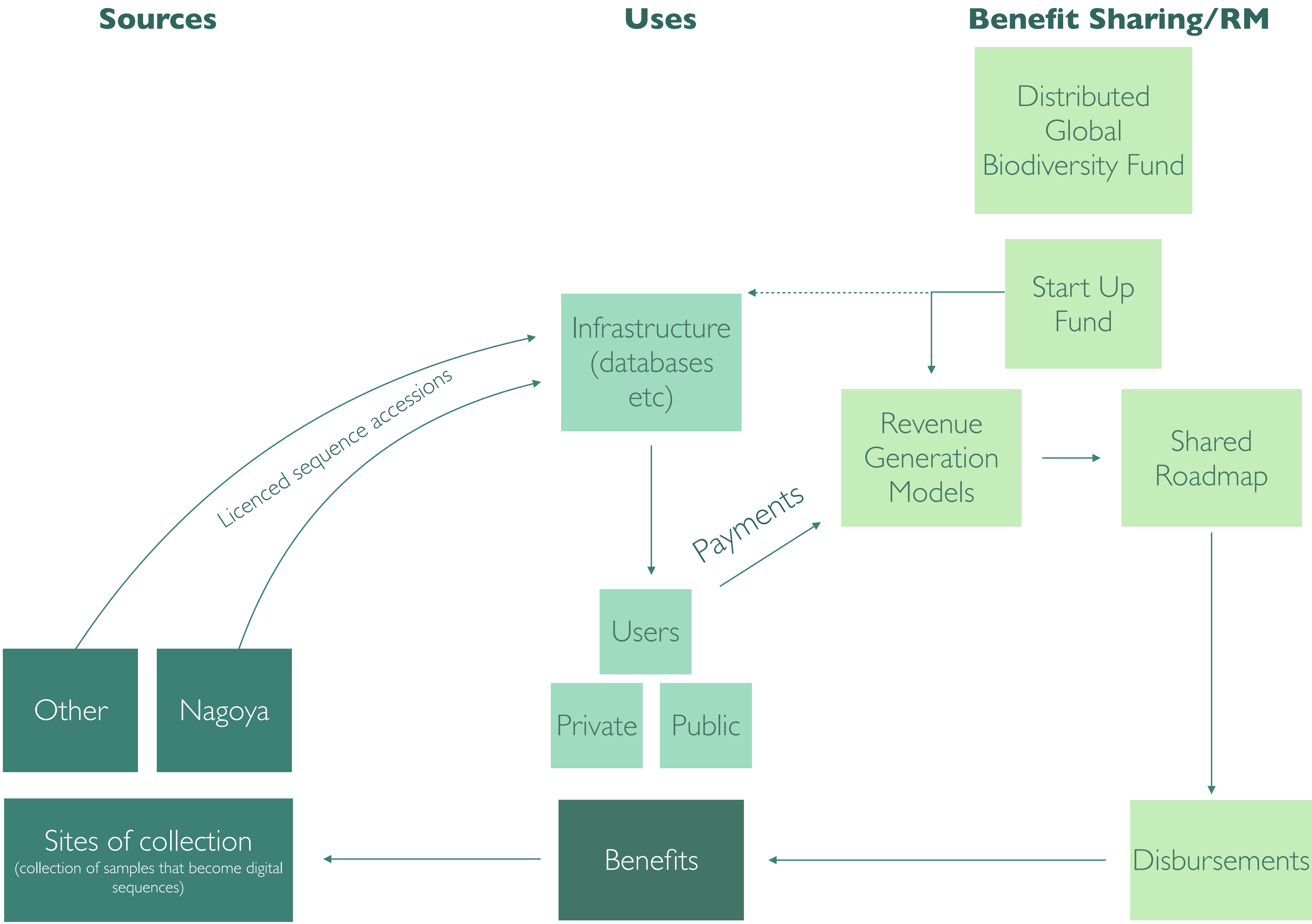
Visit this data here

2. Open Licences Work

- **Foundational Requirement:** Participating parties require that all sequences are deposited with infrastructure participating in the benefit-sharing mechanism.
- **Choices:** Participating parties agree a limited set of licences based on existing licences (GBIF set).
- **IPLCs:** In consultation with IPLCs additional options (e.g. biocultural labels & notices) will be adopted to support recognition of IPLC rights and interests.
- **Advantages:** Promotes confidence in sharing & indirect benefit-sharing. Provides legal certainty for companies. Improves transparency without hard track & trace (2.1 billion working examples). Accessions are governable objects, molecules not so much.

Participating parties = parties to one or more instruments or processes involved in the system or mechanism

Operational Model for A Cross Instrument Approach to DSI and Benefit Sharing (outline)



Use of terms: Other = Sequence data originating outside national jurisdictions (e.g. UNCLOS ABNJ/BBNJ) & from non-Parties to the Nagoya Protocol.

Source: Oldham & Kindness 2022 Sharing Digital Sequence Information. Study for the European Commission.

3. Revenue Generating Measures

- **All participating parties agree to take measures to generate revenue***
- Parties commit to a cold start fund (short term startup fund to initiate activity)
- Private sector as key contributors & partners
- **Participating parties implement one or more revenue generation model** (national/regional implementation required):
  - Cloud computing (proposed starter model)
  - Bioproduct levy (e.g. Africa Group 1%)
  - Micro levies (e.g. sales of reagents, WiLDSI paper)
  - Patent royalty rates
  - Membership fees/subscriptions/tokens/other models

* Assumes that participating parties will implement models appropriate to their economic circumstances

dsi user number assumption	user count per year	daily pp	yearly pp	yearly revenue	year five
ELIXIR 2019	2,867,265	0.1	24	68,814,360	344,071,800
NCBI 2018	4,500,000	0.1	24	108,000,000	540,000,000
GenBank (Rohden et al 2019)	5,800,000	0.1	24	139,200,000	696,000,000
INSDC median (Rohden et al 2019)	12,500,000	0.1	24	300,000,000	1,500,000,000
Global (Rohden et al 2019)	500,000,000	0.1	24	12,000,000,000	60,000,000,000
ELIXIR 2019	2,867,265	1.0	240	688,143,600	3,440,718,000
NCBI 2018	4,500,000	1.0	240	1,080,000,000	5,400,000,000
GenBank (Rohden et al 2019)	5,800,000	1.0	240	1,392,000,000	6,960,000,000
INSDC median (Rohden et al 2019)	12,500,000	1.0	240	3,000,000,000	15,000,000,000
Global (Rohden et al 2019)	500,000,000	1.0	240	120,000,000,000	600,000,000,000

Modelling revenue for use of the infrastructure using cloud storage subscriptions (Apple icloud, Google Photos)

US$ 24 dollars a year per person

US$ 240 dollars a year per person

Potential revenue is an outcome of the number of users x price point (subscription)

Why A Distributed Global Biodiversity Fund?

- Resource mobilisation will require revenue generation measures implemented nationally/regionally rather than a “donor” model;
- The biodiversity and life science infrastructure is globally distributed but fragile in terms of funding. Support best addressed on the national/regional level;
- Recommended GEF-8 replenishment US$1.6-3.7 billion (2022-2026) (CBD/SBI/3/5/Add.2/Rev.1); Annual biodiversity funding gap is US$103-178 billion (low) and US$599-823 billion (high) (CBD/SBI/3/5/Add.2/Rev.1);
- The GEF is not equipped to handle that potential volume of funds (World Bank Group AUM = US$ 70 billion, 2021 98 billion commitments, 60 billion disbursements);
- Some countries have already created national biodiversity funds.*

*India 2002 National Biodiversity Act. Germany 2021 US$1 billion Legacy Landscapes Fund. China 2021 US$232 million biodiversity fund. UK 2021 £100 million Biodiversity Landscapes Fund, Japan 2010 Biodiversity Fund, Brazil 1996 FUNBIO, Caribbean Biodiversity Fund 2012 cerca US$102 million.

5. Shared Roadmap

- A shared road map seeks to operationalise the post 2020 GBF. It could focus on operational investments in:
  - Priorities identified by developing countries and IPLCs (more bang per buck)*
  - The sustainability of the biodiversity and life science infrastructure
  - Capacity building
  - Piloting innovative revenue generation measures between participating parties
  - Facilitating match making between parties with shared interests in implementation (e.g. on IPLCs)

* Low income countries have the most to gain from increased investment in conservation with lower opportunity costs.
Only 13% of total biodiversity spending & 5% of total conservation spending is allocated to developing countries. See CBD/SBI/3/5/Add.2/Rev.1 para 67 (f) & para 69.

6. Participants

- **The private sector** is a primary focus of this proposal
  - As partners in service provision (e.g. cloud computing)
  - As paying users (e.g. of infrastructure, levies etc)
  - In helping design innovative revenue models
- **Research funding bodies & IPBES** are important partners for:
  - Coordinating shared research agendas (e.g. on neglected areas of research such as the water column in the oceans, taxonomy, neglected foods & diseases)
- **The Biodiversity and Life Science Infrastructure**
- **IPLCs** as the main stewards of biodiversity
- **Technical Assistance Agencies** as match making, implementing and reporting agencies

Conclusion

- The proposed cross instrument approach would:

  a) Provide a scaleable global solution to sharing digital sequence information using open licences;

  b) Directly link digital sequence information to resource mobilisation;

  c) Recognise the need to support the biodiversity & life science infrastructure;

  d) Directly engage the private sector as contributors & partners;

  e) Use existing administrative and accounting infrastructure;

  e) Avoid costly new global instruments and bodies.