



proteus

Data Repository

Joe Turner, Programme Officer, UNEP-WCMC

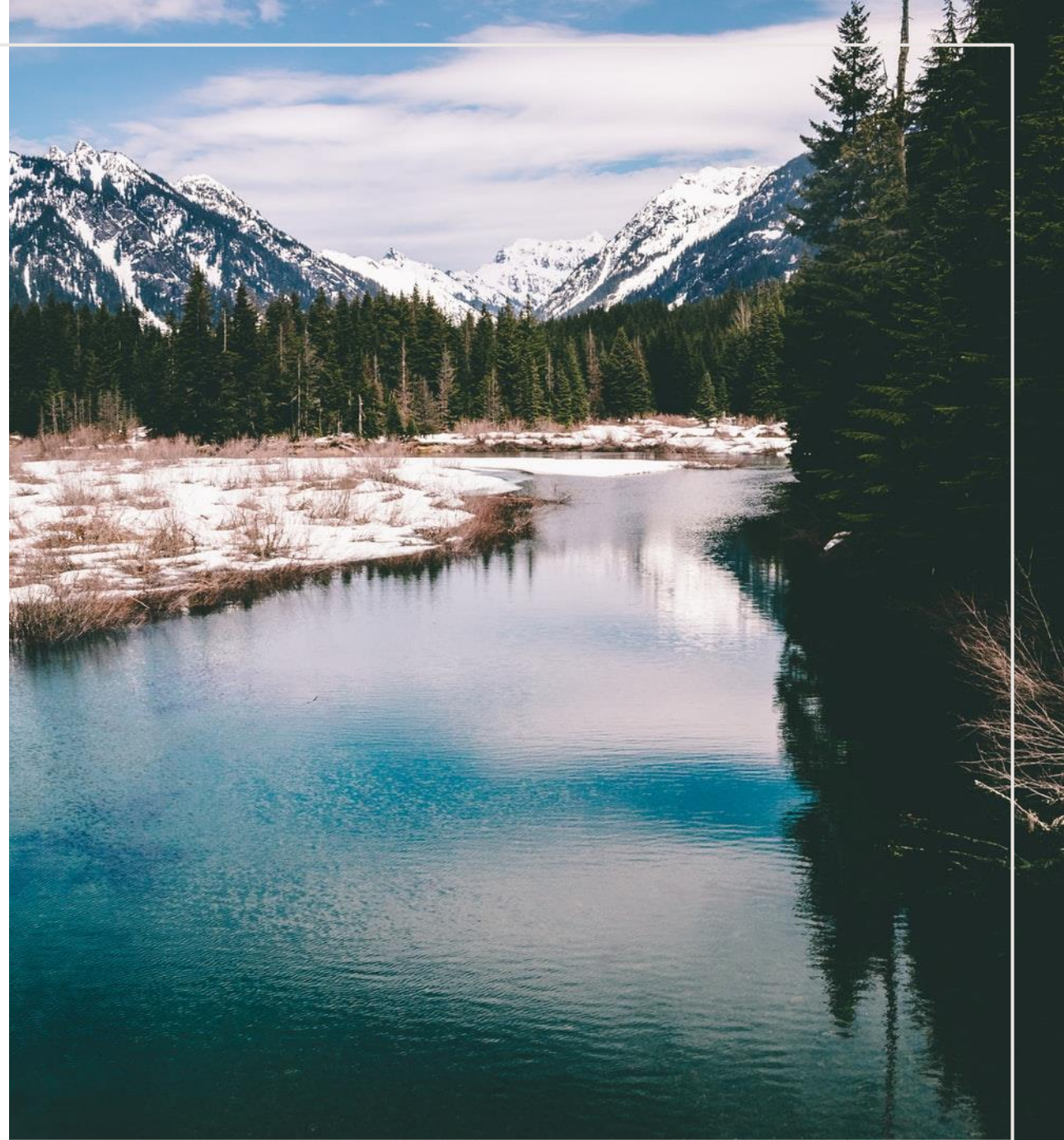
Proteus Annual Meeting

22/11/2021

proteus GOAL 2

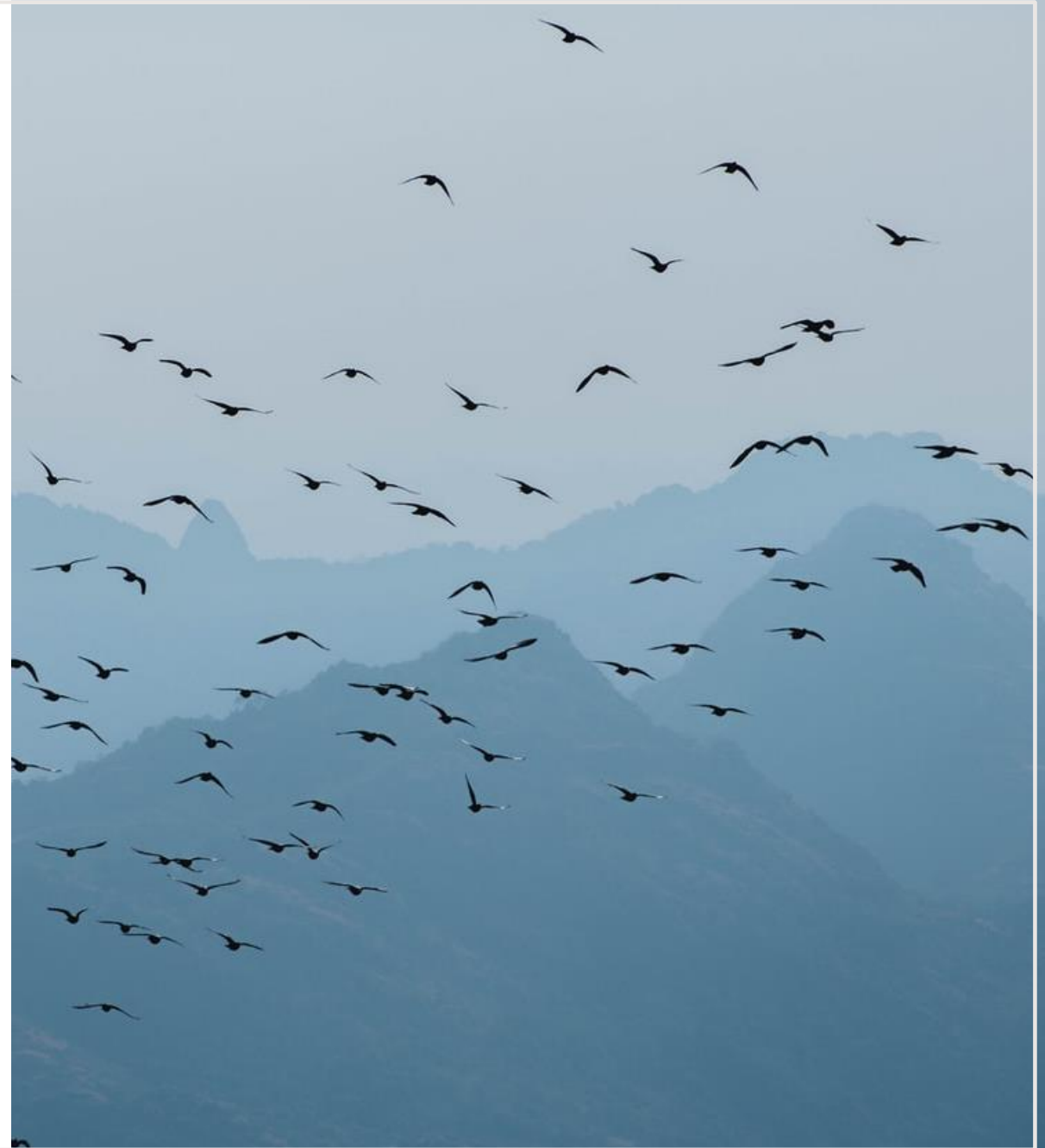
Data sharing

- This phase of Proteus will explore mechanisms to enable data sharing
- This will allow observation data collected by the private sector to feed into global datasets
- This phase of Proteus will lay the groundwork for as much data as possible to be made freely available



WHY SHARE DATA?

- Sharing open data can increase the visibility and impact of institutions.
- Transparency.
- Support reporting commitments.
- Contribute to global knowledge about biodiversity.



COLLABORATION WITH GBIF

- UNEP-WCMC GBIF Node.

A team designated by a Participant to coordinate a network of people and institutions that produce, manage and use biodiversity data, collectively building an infrastructure for delivering biodiversity information.

- Improve quality and availability of biodiversity data.
- Create partnerships and alignment of efforts.
- Enables the integration of data.
- Raise the visibility of data.



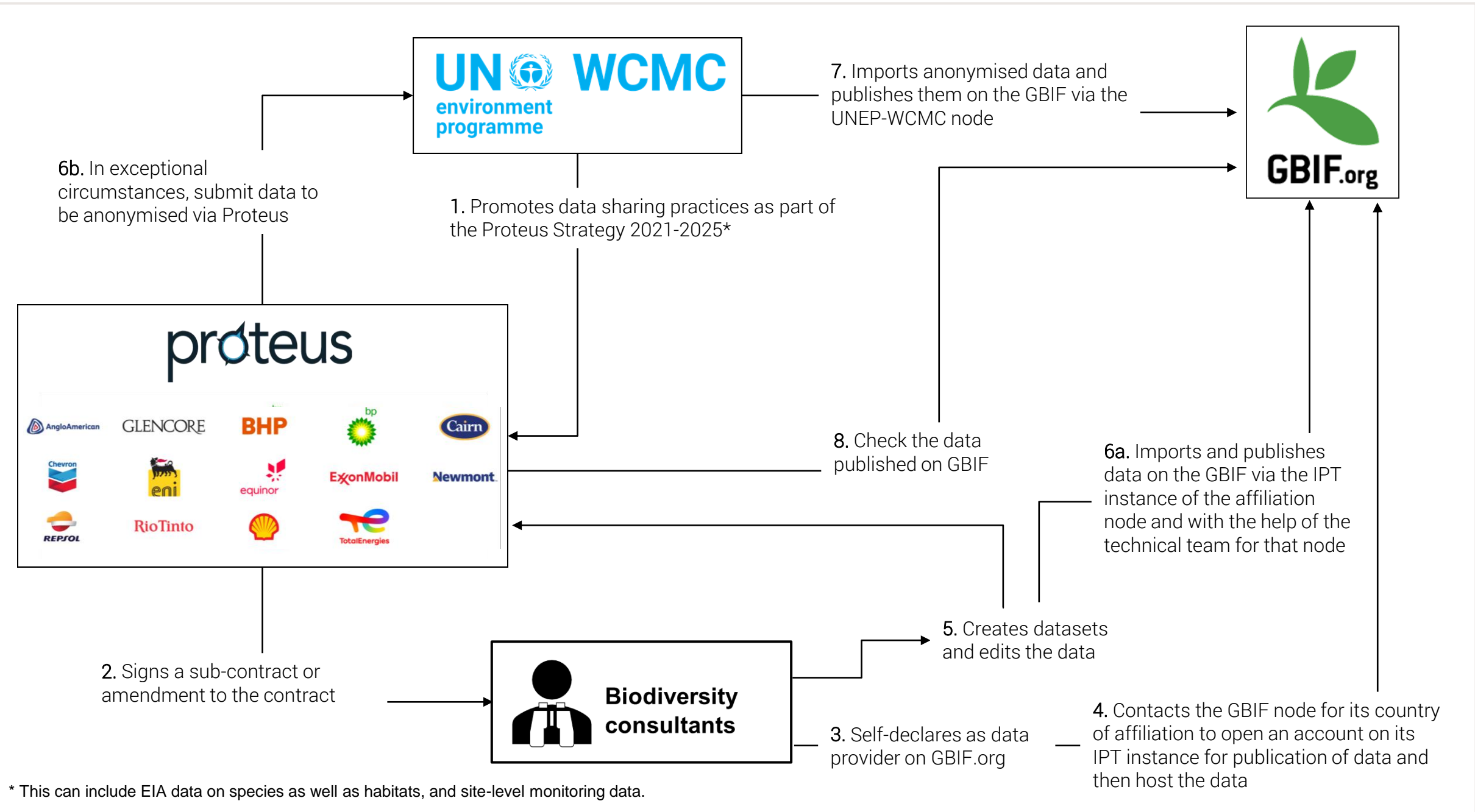
DATA FORMATTING

- Darwin core format required for data submission.
- [Generalise Sensitive Species Occurrence Data](#).
- Select license (CC0, CC BY, CC BY-NC).

Darwin Core records

Term	Status
eventID	Required
eventDate	Required
samplingProtocol	Required
samplingSizeValue & samplingSizeUnit	Required
countryCode	Strongly recommended
parentEventID	Strongly recommended
samplingEffort	Strongly recommended
locationID	Strongly recommended
decimalLatitude & decimalLongitude2	Strongly recommended
geodeticDatum	Strongly recommended
coordinateUncertaintyInMeters	Strongly recommended
footprintWKT	Strongly recommended
occurrenceStatus	Strongly recommended





* This can include EIA data on species as well as habitats, and site-level monitoring data.

DATA EXAMPLE

The screenshot shows a web interface for a species profile. At the top, a green navigation bar contains links for 'Get data', 'How-to', 'Tools', 'Community', and 'About', along with utility icons for home, search, and a message box. The main content area is titled 'SPECIES | ACCEPTED' and features the species name *Pongo pygmaeus* (Linnaeus, 1760) in a large font. Below the name, it lists 'Published in: Amoenit. Acad. vol.6 p.68' and 'source: The Integrated Taxonomic Information System'. A 'Bornean Orangutan' label is present in English, and the 'Basionym' is listed as *Simia pygmaeus* Linnaeus, 1760. A secondary navigation bar includes 'OVERVIEW', '1 TREATMENT', 'METRICS', and 'REFERENCE TAXON', with '5,262 OCCURRENCES' and '8 INFRASPECIES' highlighted in green. The '433 OCCURRENCES WITH IMAGES' section displays a horizontal gallery of six photographs of orangutans in their natural habitat, with a 'SEE GALLERY' button on the right. Below this, the '573 GEOREFERENCED RECORDS' section features a world map with yellow dots indicating the species' distribution, primarily in Southeast Asia. The map includes zoom controls and a 'Generated 2 hours ago' timestamp. At the bottom, there is a 'Any year' filter set to '1810 - 2020' and an 'EXPLORE' button with additional icons.

Classification

Select a species

Kingdom **Animalia**

Phylum **Chordata**

Class **Mammalia**

Order **Primates**

Family **Hominidae**

Genus *Pongo* Lacépède, 1799

Species *Pongo pygmaeus* (Linnaeus, 1760)

≡ *Simia pygmaeus* Linnaeus, 1760

Immediate children

Subspecies *Pongo pygmaeus* subsp. *ciochoni* Schwartz et al., 1995

Subspecies *Pongo pygmaeus* subsp. *devosi* Schwartz et al., 1995

Subspecies *Pongo pygmaeus* subsp. *fromageti* Schwartz et al., 1995

Subspecies *Pongo pygmaeus* subsp. *kahlkei* Schwartz et al., 1995

Subspecies *Pongo pygmaeus* subsp. *morio* (Owen, 1837)

Subspecies *Pongo pygmaeus* subsp. *pygmaeus*

Subspecies *Pongo pygmaeus* subsp. *wurmbii* (Tiedemann, 1808)

Unranked **BOLD:AAC3724** (cf. *Pongo pygmaeus*)

SPECIES | ACCEPTED

Pongo pygmaeus (Linnaeus, 1760)

Published in: Amoenit. Acad. vol.6 p.68 source: The Integrated Taxonomic Information System

Bornean Orangutan In English **Basionym:** *Simia pygmaeus* Linnaeus, 1760

OVERVIEW 1 TREATMENT METRICS REFERENCE TAXON

5,262 OCCURRENCES 8 INFRASPECIES

433 OCCURRENCES WITH IMAGES

573 GEOREFERENCED RECORDS

Generated 2 hours ago © OpenStreetMap contributors, © OpenMapTiles, © GBIF.

Any year 1810 - 2020 EXPLORE

DATA EXAMPLE

The screenshot displays a web interface for searching biodiversity data. The top navigation bar includes links for 'Get data', 'How-to', 'Tools', 'Community', and 'About'. A search bar on the right contains the text 'jt3ksc'. The main content area is titled 'SEARCH OCCURRENCES | 280 WITH COORDINATES'. Below this, there are tabs for 'TABLE', 'GALLERY', 'MAP', 'TAXONOMY', 'METRICS', and 'DOWNLOAD', with 'MAP' currently selected. The map shows a world view with orange dots representing occurrences, primarily clustered in Southeast Asia. On the left, a sidebar titled 'Occurrences' contains a search bar and a list of filters:

- Occurrence status:** Present
- Licence:** (empty)
- Scientific name:** Pongo pygmaeus (Linnaeus, 1760)
- Verbatim scientific name:** (empty)
- Basis of record:**
 - Observation (1)
 - Machine observation (157)
 - Human observation (280)
 - Material sample (0)
 - Literature (0)
 - Preserved specimen (125)
 - Fossil specimen (10)
 - Living specimen (0)
 - Unknown (0)
- Location:** Including coordinates
- Administrative areas (gadm.org):** (empty)
- Coordinate uncertainty in metres:** (empty)
- Year:** (empty)
- Month:** (empty)
- Dataset:** (empty)
- Country or area:** (empty)

Buttons for 'CLEAR' and 'REVERSE' are located below the 'Basis of record' section. The bottom of the page includes a copyright notice: '© OpenStreetMap contributors, © OpenMapTiles, GBIF'.

DATA EXAMPLE

Occurrences 4

Scientific name ▼

Pongo pygmaeus (Linnaeus, 1760)

Verbatim scientific name ▼

Basis of record ^

<input type="checkbox"/> Observation	1
<input type="checkbox"/> Machine observation	157
<input checked="" type="checkbox"/> Human observation	280
<input type="checkbox"/> Material sample	0
<input type="checkbox"/> Literature	0
<input type="checkbox"/> Preserved specimen	125
<input type="checkbox"/> Fossil specimen	10
<input type="checkbox"/> Living specimen	0
<input type="checkbox"/> Unknown	0

[CLEAR](#) [REVERSE](#)

Location ▼

Including coordinates

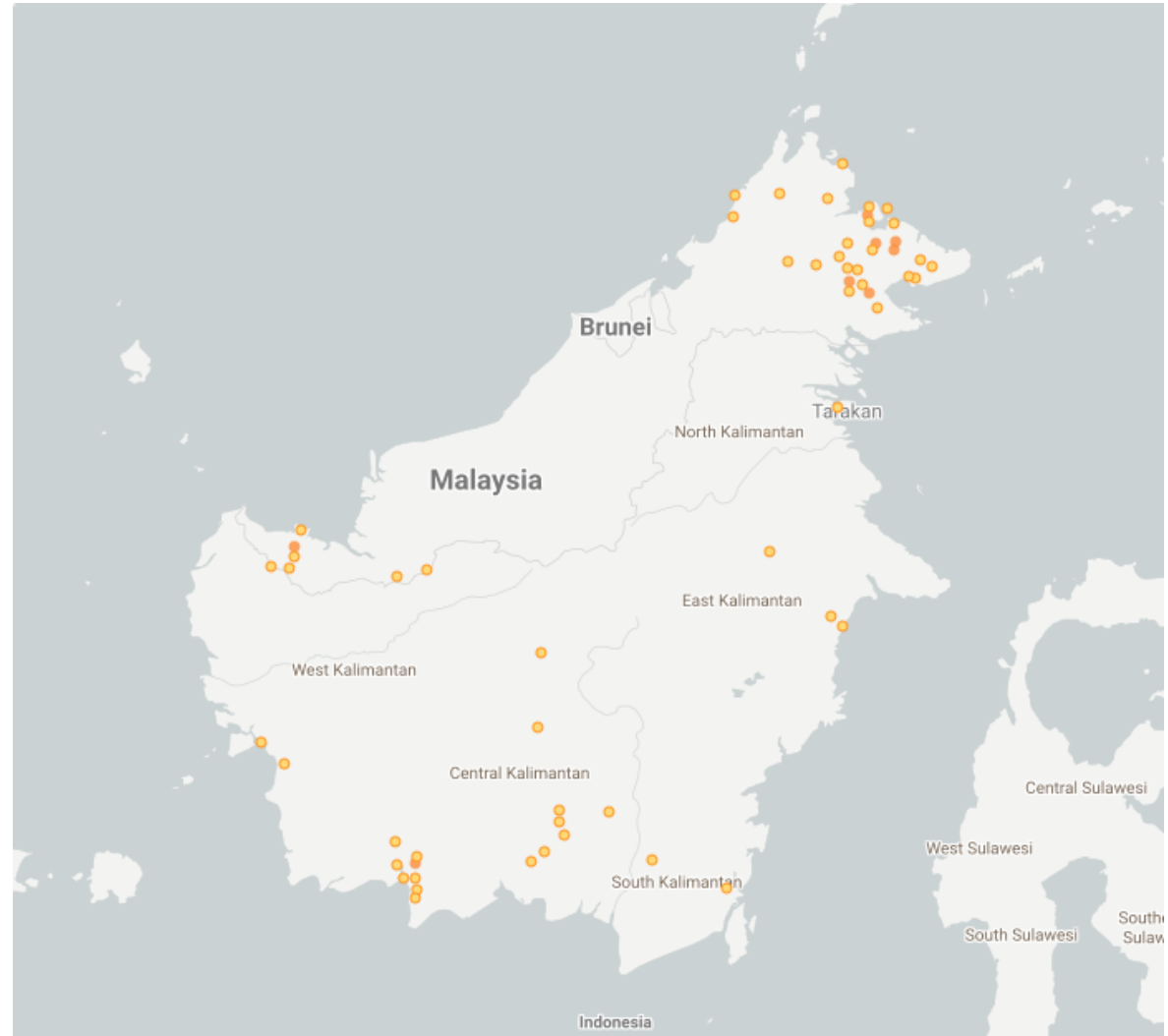
Administrative areas (gadm.org) ▼

Coordinate uncertainty in metres ▼

Year ^

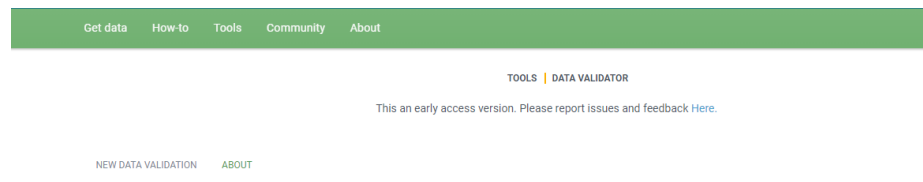
No filter ×

1000 2021



TRAINING AND RESOURCES

- [Training and Resources.](#)
- Tools – e.g. [Data Validator.](#)



What is the GBIF data validator?

The GBIF data validator is a service that allows anyone with a GBIF-relevant dataset to receive a report on the syntactical correctness and the validity of the content contained within the dataset. By submitting a dataset to the validator, you can go through the validation and interpretation procedures usually associated with publishing in GBIF and quickly determine potential issues in data - without having to publish it.

How does it work?

You start by uploading the dataset file to the validator, either by 1) clicking *SELECT FILE* and selecting it on your local computer or 2) dragging the file from a local folder and dropping it on the *Drop here* icon. You can also enter the URL of a dataset file accessible from the internet. This is particularly useful for larger datasets. Once you hit the *Submit* button, the validator starts processing your dataset file. You will be taken straight to a page showing the status of the validation.

Depending on the size of your dataset, processing might take a while. You don't have to keep the browser window open, as a unique job ID is issued every time a new validation process is started. If your dataset is taking too long to process, just save the ID (bookmark the URL) and use it to return at a later time to view the report. We'll keep the report for a month, during which you can come back whenever you like.



Training and e-Learning

GBIF empowers its global network and strengthens its community of practice through training activities that enable members to mobilize and use biodiversity information effectively.



Serge Roberteau Tchotto (BIOPAMA, left) and Elias Bizatu (National University of Rwanda, right) at the BID African regional meeting and nodes training, 3 Sept 2019, Yaounde, Cameroon. Photo by Malliane Raymond | GBIF, licensed under CC BY 4.0.

The GBIF Secretariat works closely with its volunteer community of trainers and mentors to develop and maintain several online training courses. These courses are conducted either as online or on-site workshops.



Thank You



proteus

Data Repository

Joe Turner, Programme Officer, UNEP-WCMC

Proteus Annual Meeting

22/11/2021

UN 
**environment
programme**

WCMC