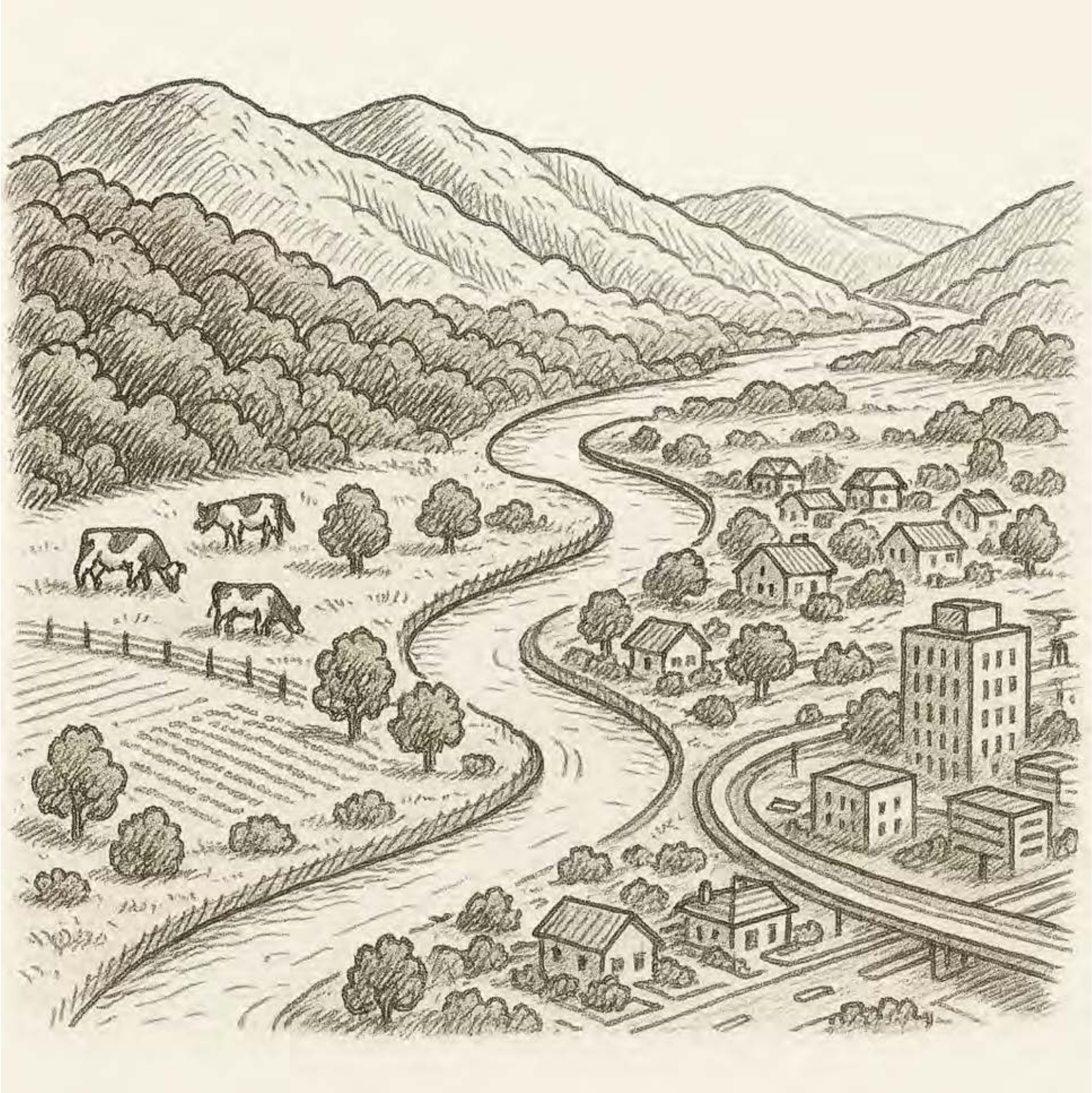
# eDNA bridge to GBIF

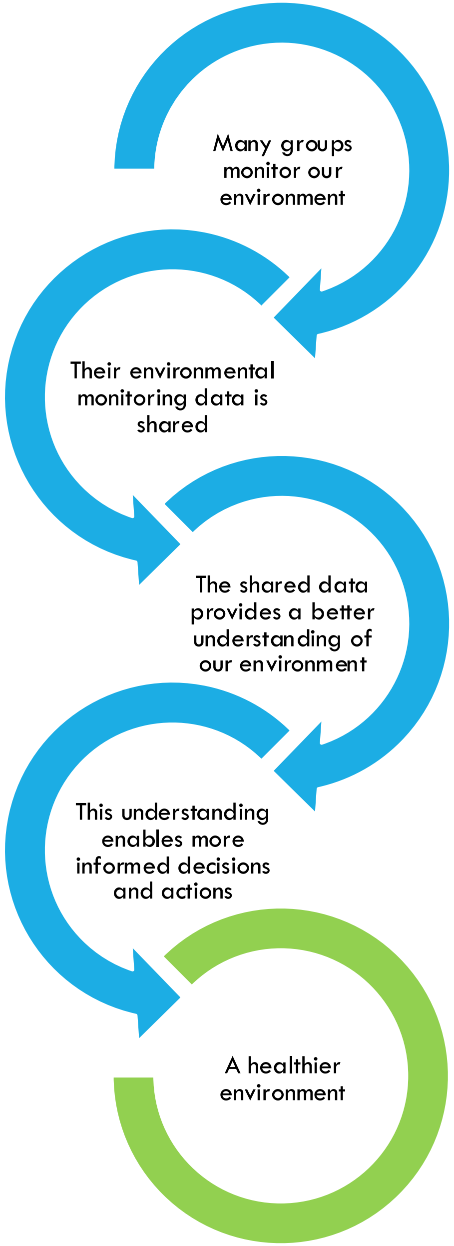
Fiona Hodge, Carolyn Mander and Georgia Sharpe

Ministry for the
Environment
Manatū Mō Te Taiao

# We all want a healthier environment

Ministry for the Environment
Manatū Mō Te Taiao





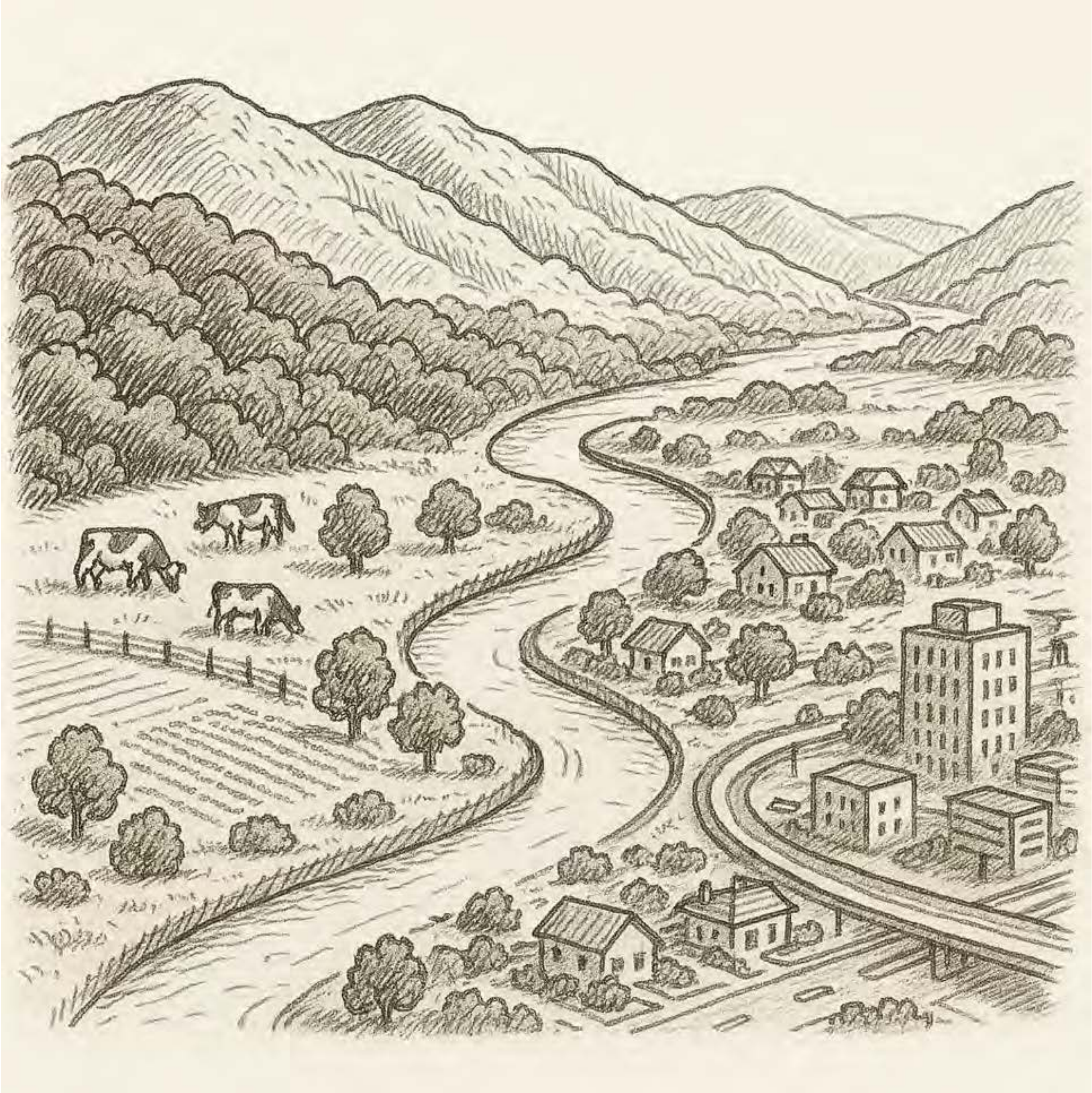
Many groups monitor our environment

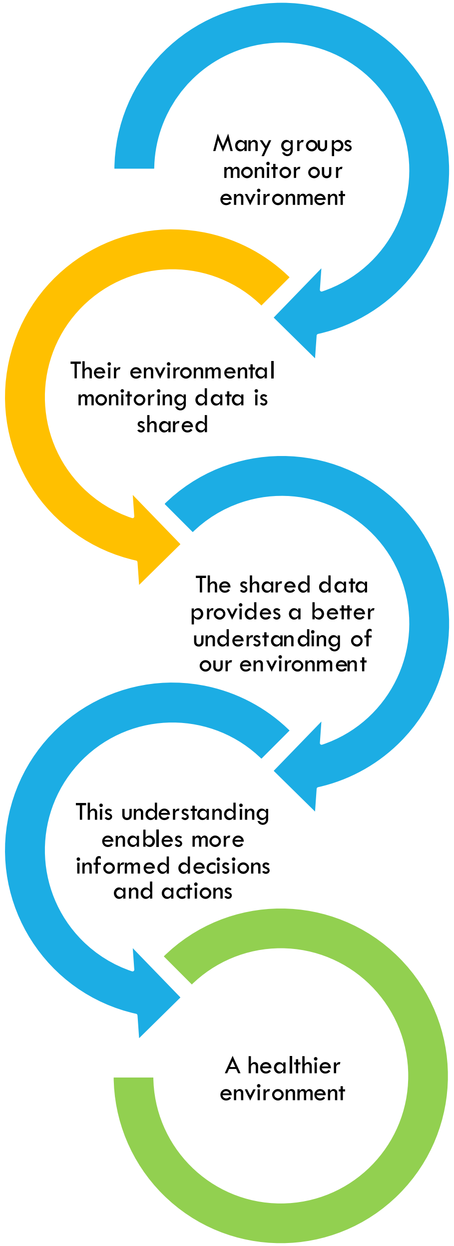
Their environmental monitoring data is shared

The shared data provides a better understanding of our environment

This understanding enables more informed decisions and actions

A healthier environment





Many groups monitor our environment

Their environmental monitoring data is shared

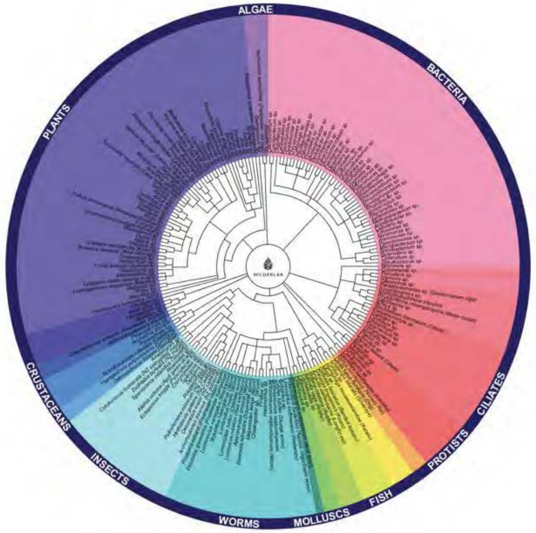
The shared data provides a better understanding of our environment

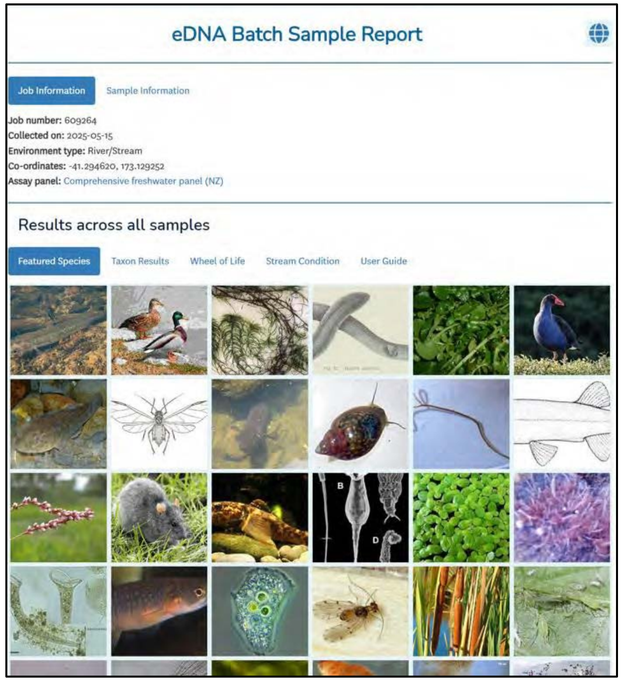
This understanding enables more informed decisions and actions

A healthier environment

# Environmental DNA (eDNA)

*Ministry for the*
**Environment**
*Manatū Mō Te Taiao*




eDNA Batch Sample Report

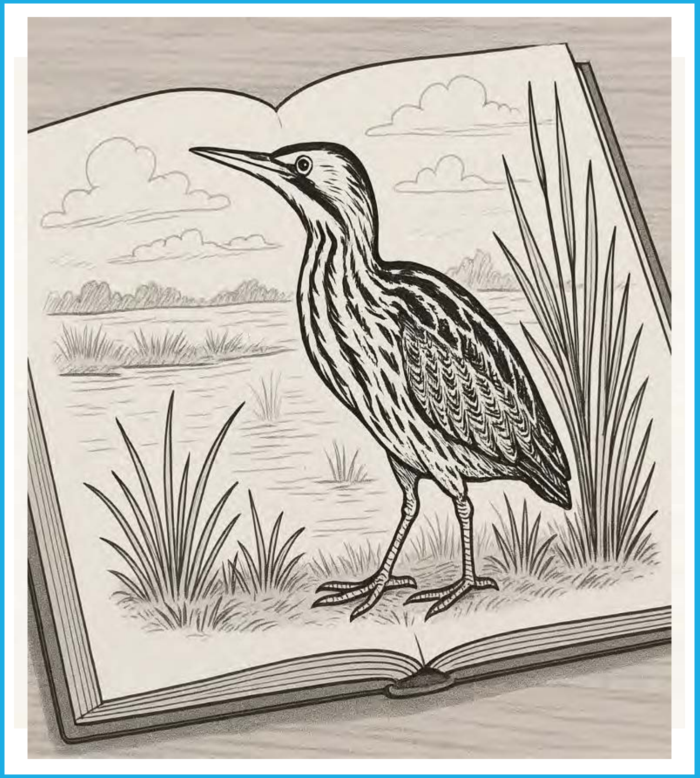
Job Information | Sample Information

**Job number:** 609264
**Collected on:** 2025-05-15
**Environment type:** River/Stream
**Co-ordinates:** -41.294620, 173.129252
**Assay panel:** Comprehensive freshwater panel (NZ)

Results across all samples

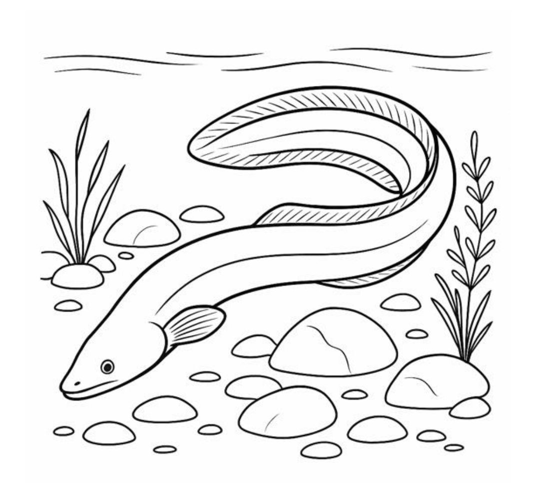
Featured Species | Taxon Results | Wheel of Life | Stream Condition | User Guide

# Each eDNA sample holds many stories

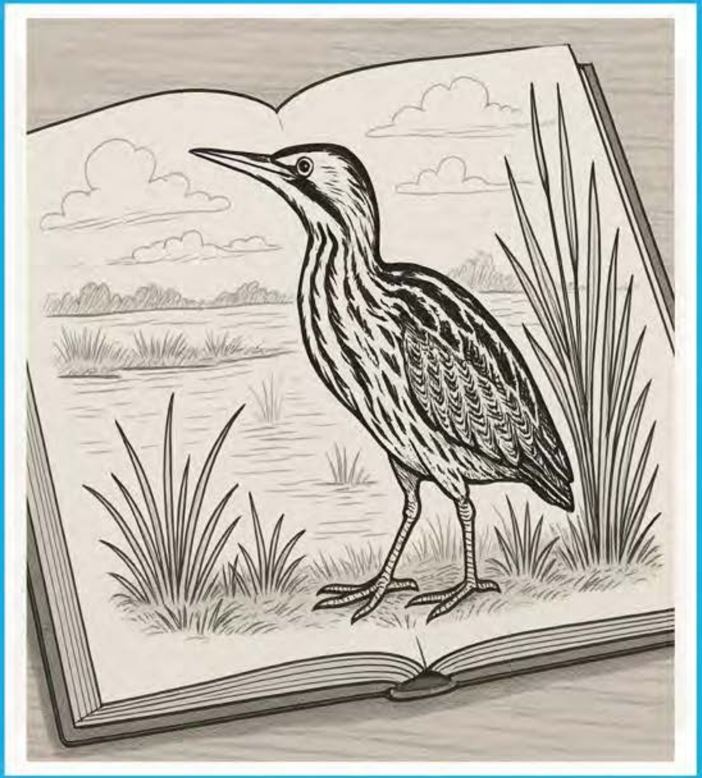


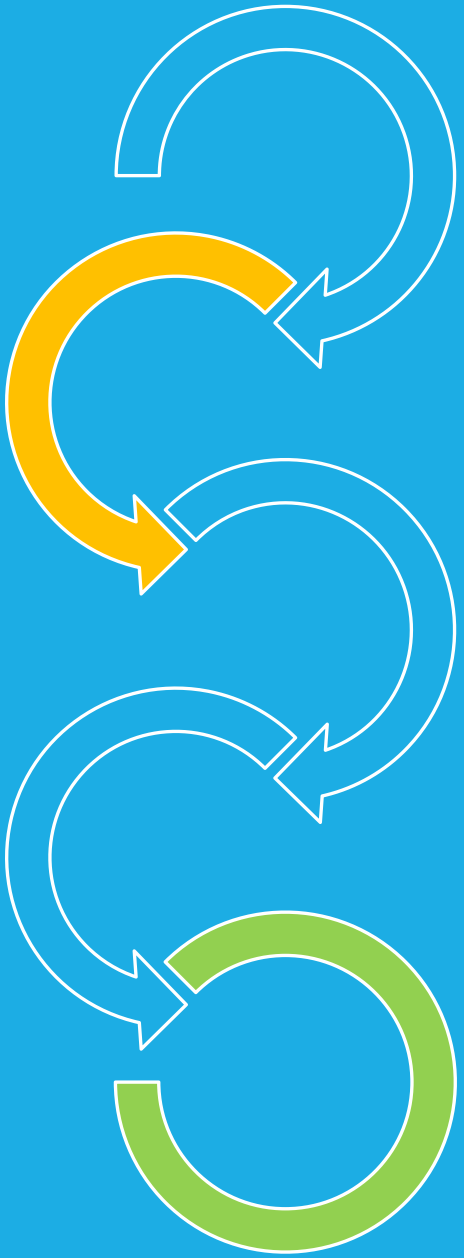
# Each eDNA sample holds many stories

# Each eDNA sample holds many stories



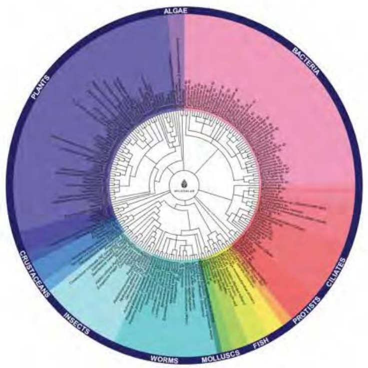
If we don’t share, we miss out on crucial stories





# Our starting point

WILDERLAB




eDNA Batch Sample Report

**Job number:** 609264
**Collected on:** 2025-05-15
**Environment type:** River/Stream
**Co-ordinates:** -41.294820, 173.123252
**Assay panel:** Comprehensive freshwater panel (NZ)

Results across all samples

GBIF | Global Biodiversity Information Facility

Datasets

**112,514**

Hosted portals

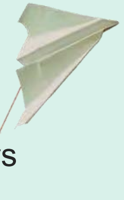
**53**

Country Participants

**65**

Peer-review papers using data

**12,636**



Organizational Participants

**42**

Average records downloaded per month (2025)

**267.6 billion**

Publishers

**2,427**

Species occurrence records

**3,089,709,484**

By the numbers | 1 April 2025

www.gbif.org

# Building a bridge

EPI-interactive

▼

epi

Ministry for the Environment
Manatū Mō Te Taiao

# Building a bridge

Ministry for the
Environment
Manatū Mō Te Taiao

May 2026

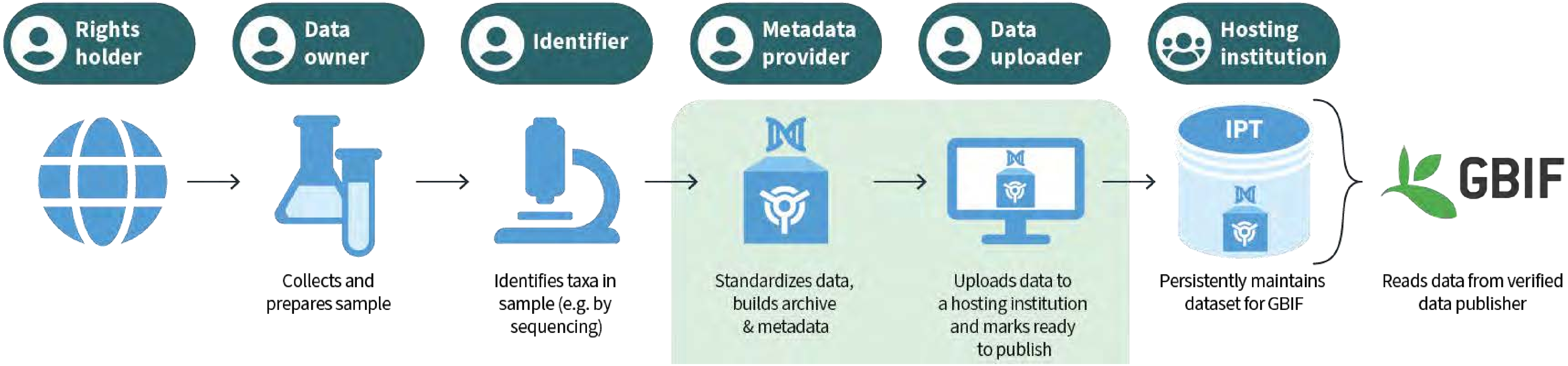
# Streamlining the upload of biodiversity data to global repositories with *eDNAbridge*

Kiel Hards – Senior Data Scientist, Epi

**R Exchange 2026**

# How do we balance bulk eDNA submission with data integrity and quality validation?

Image source: https://fishbio.com/traces-left-behind/



**Rights holder**

**Data owner**
Collects and prepares sample

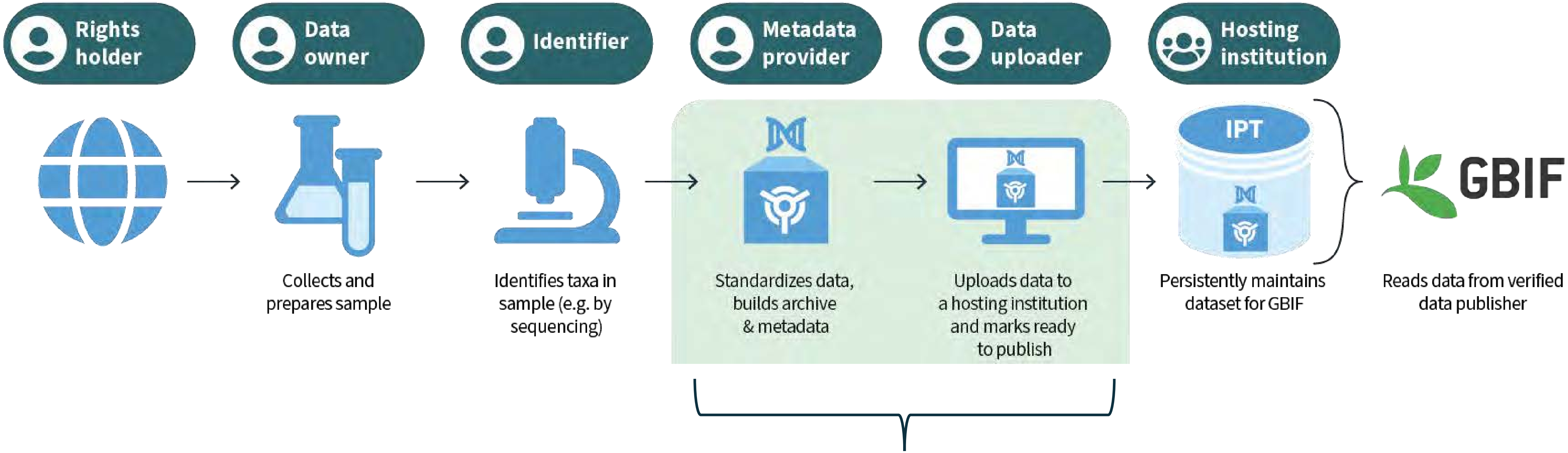
**Identifier**
Identifies taxa in sample (e.g. by sequencing)

**Metadata provider**
Standardizes data, builds archive & metadata

**Data uploader**
Uploads data to a hosting institution and marks ready to publish

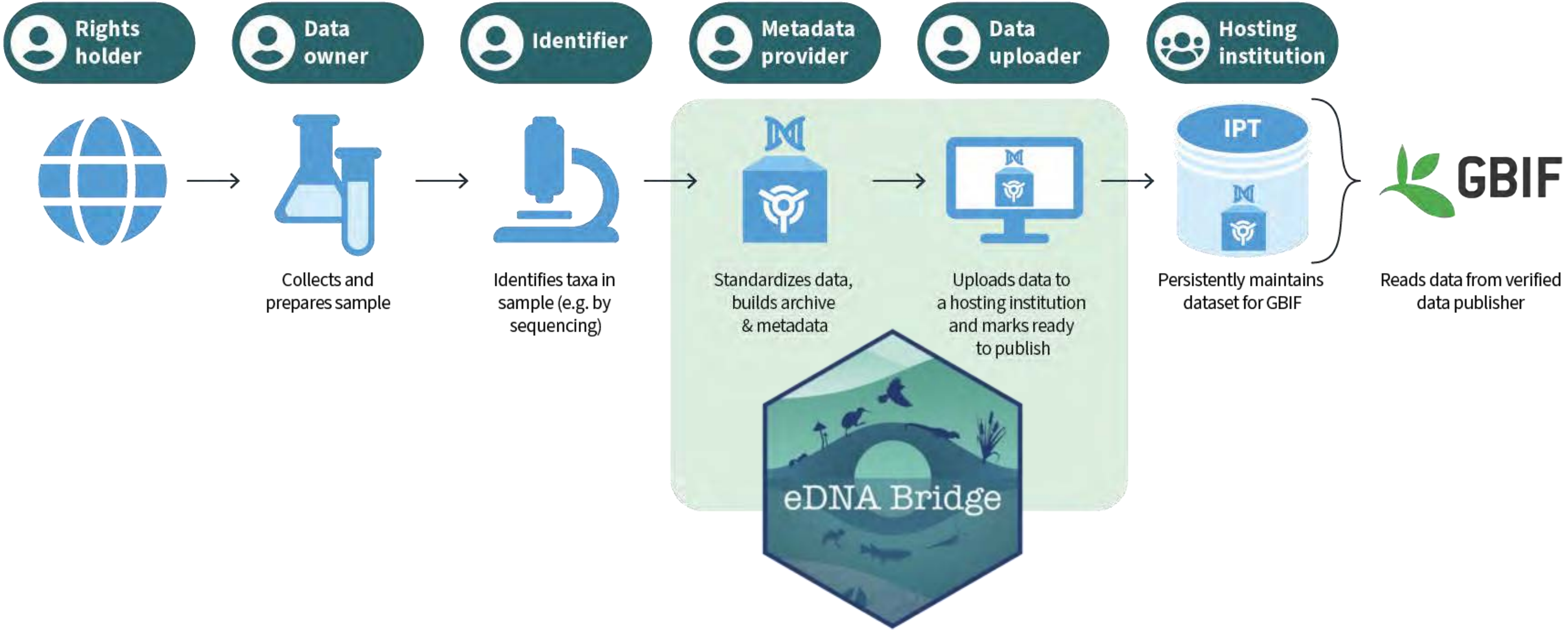
**Hosting institution**
Persistently maintains dataset for GBIF

Reads data from verified data publisher



- **Rights holder**
- **Data owner**: Collects and prepares sample
- **Identifier**: Identifies taxa in sample (e.g. by sequencing)
- **Metadata provider**: Standardizes data, builds archive & metadata
- **Data uploader**: Uploads data to a hosting institution and marks ready to publish
- **Hosting institution**: Persistently maintains dataset for GBIF
- Reads data from verified data publisher

Technically Challenging
Time Consuming



**Rights holder**

**Data owner**

Collects and prepares sample

**Identifier**

Identifies taxa in sample (e.g. by sequencing)

**Metadata provider**

Standardizes data, builds archive & metadata

**Data uploader**

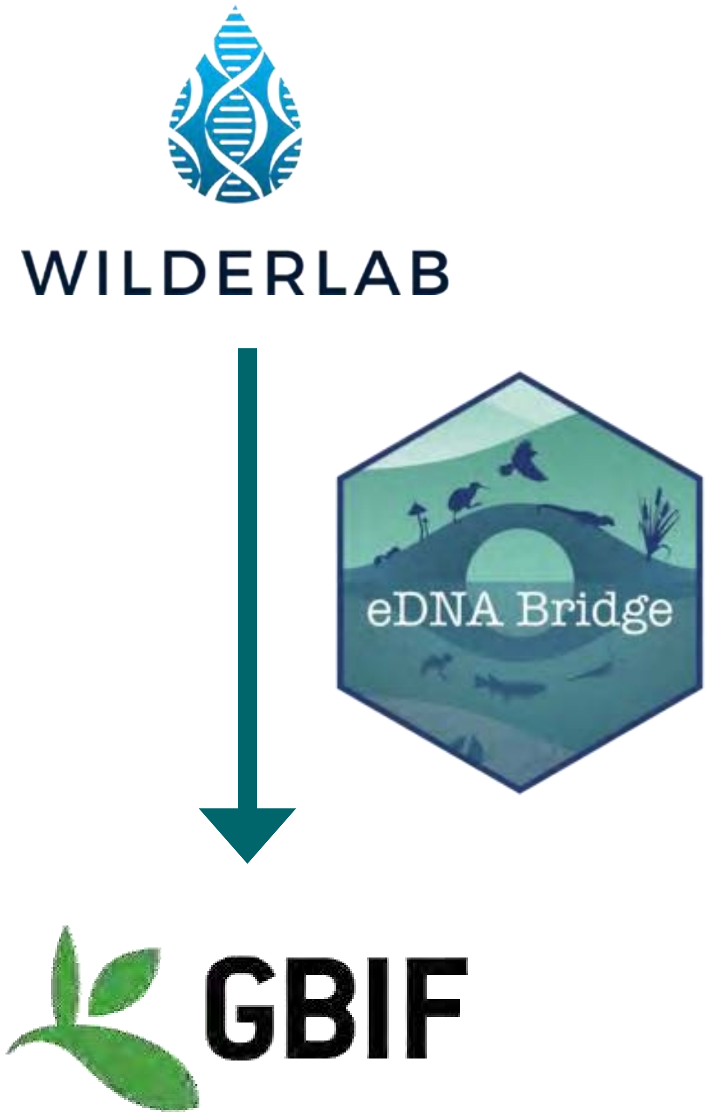
Uploads data to a hosting institution and marks ready to publish

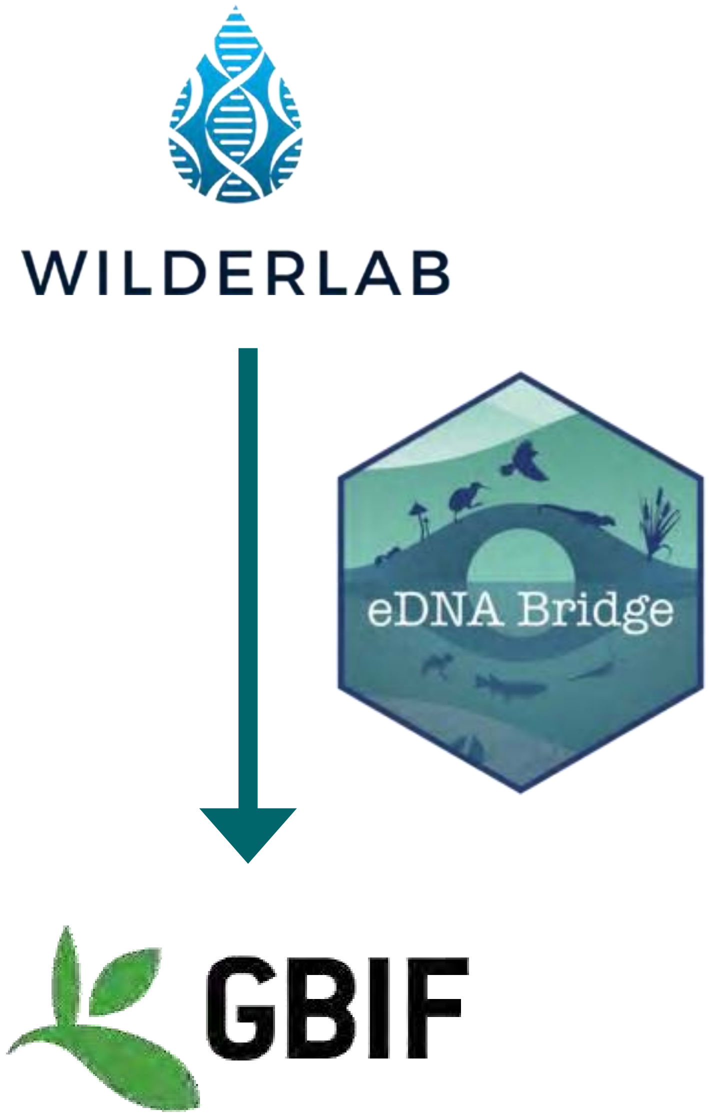
**Hosting institution**

Persistently maintains dataset for GBIF

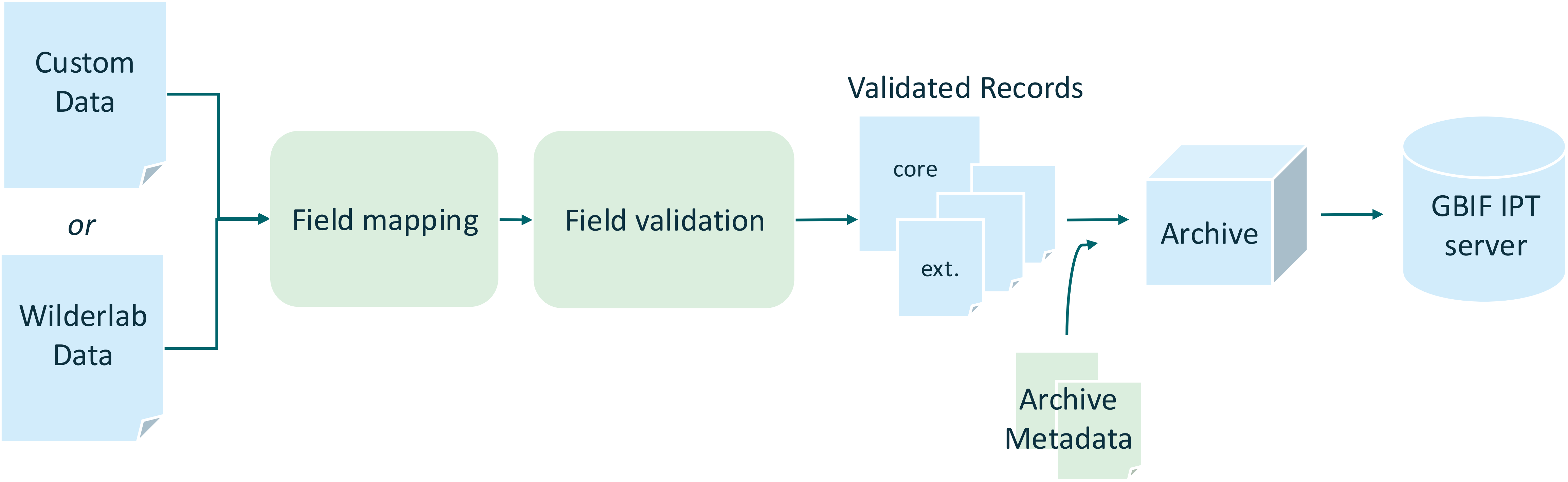
Reads data from verified data publisher

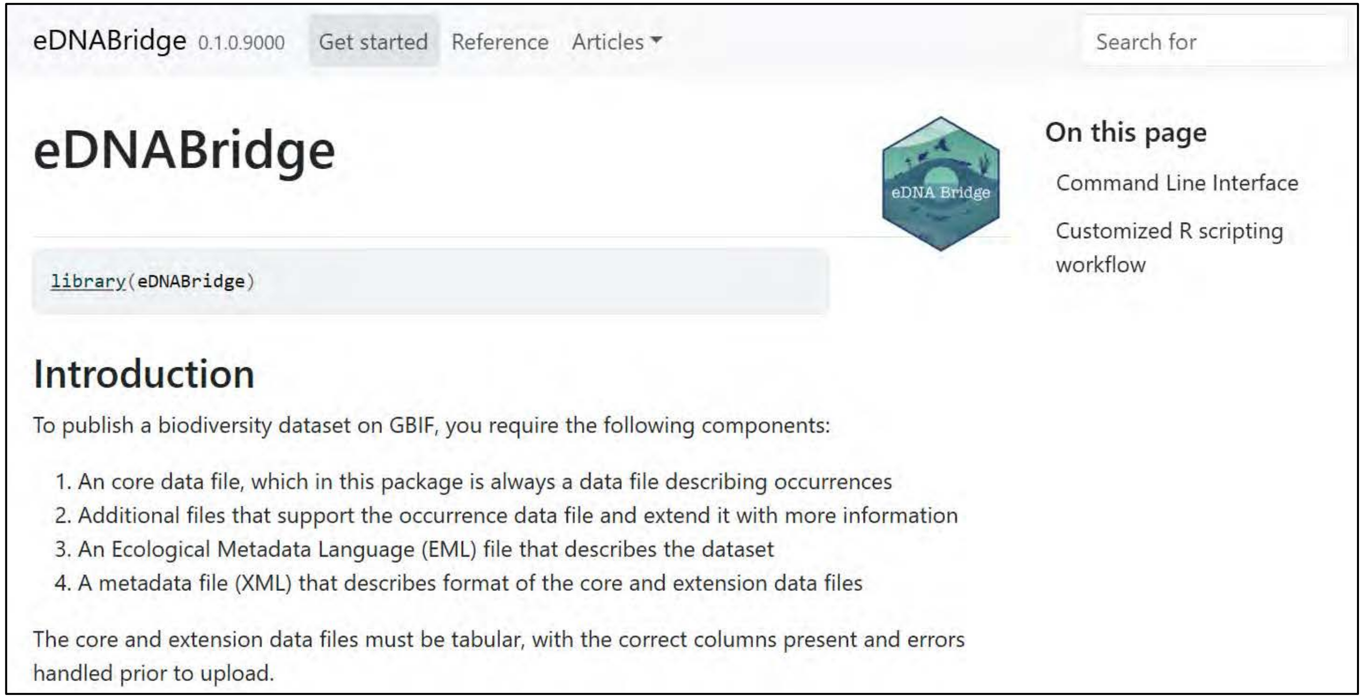
Open source approach allowed us to handle complex scientific data while staying flexible





- Open-source R package
- From sequencing laboratory outputs to publication in GBIF repository
- Aimed at bulk / automated data uploads
- Supports both guided, interactive use and integration into laboratory pipelines


Custom Data
or
Wilderlab Data
Field mapping
Field validation
Validated Records
core
ext.
Archive Metadata
Archive
GBIF IPT server



## eDNABridge

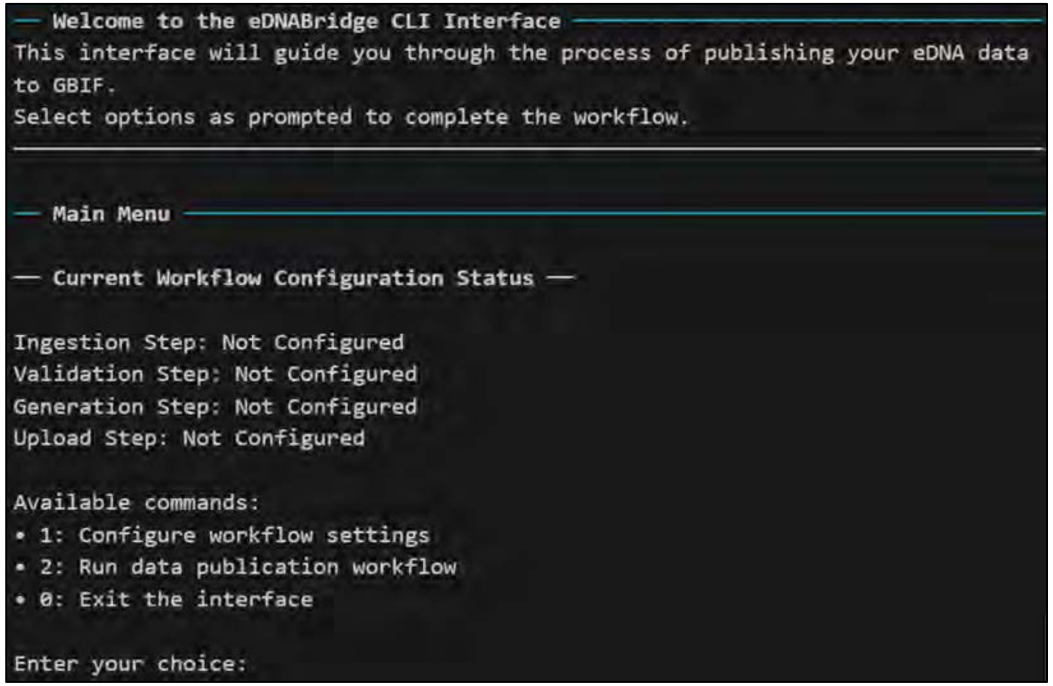
```r
library(eDNABridge)
```

### Introduction

To publish a biodiversity dataset on GBIF, you require the following components:

1. An core data file, which in this package is always a data file describing occurrences
2. Additional files that support the occurrence data file and extend it with more information
3. An Ecological Metadata Language (EML) file that describes the dataset
4. A metadata file (XML) that describes format of the core and extension data files

The core and extension data files must be tabular, with the correct columns present and errors handled prior to upload.



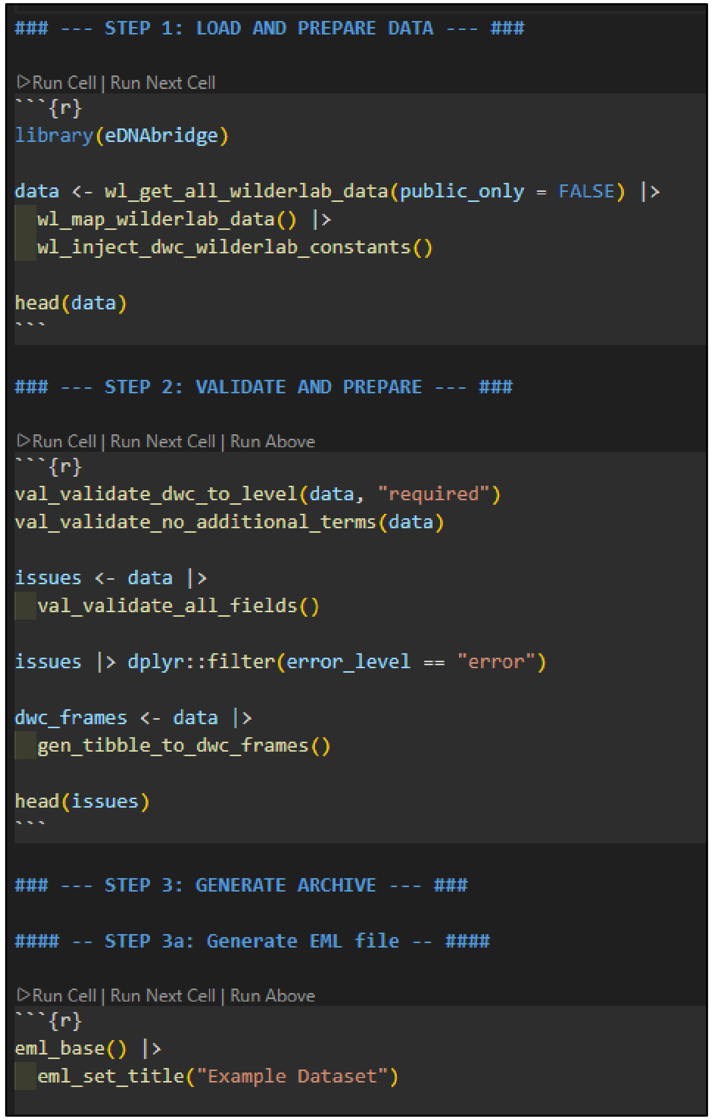
```
── Welcome to the eDNABridge CLI Interface ──────────────────────────────────
This interface will guide you through the process of publishing your eDNA data
to GBIF.
Select options as prompted to complete the workflow.
────────────────────────────────────────────────────────────────────────────


── Main Menu ────────────────────────────────────────────────────────────────

── Current Workflow Configuration Status ──

Ingestion Step: Not Configured
Validation Step: Not Configured
Generation Step: Not Configured
Upload Step: Not Configured

Available commands:
• 1: Configure workflow settings
• 2: Run data publication workflow
• 0: Exit the interface

Enter your choice:
```

# Tailored documentation, guided workflows, CLI option

# Using *eDNAbridge*

- Documentation in a pkgdown site
  - Run `pkgdown::build_site()` to view
- Run `eDNABridge()` for a CLI helper to do one-off submissions
- Multiple functions to customize workflows via extendible modules:
  - Data Ingestion & Validation
  - Metadata and Archive Creation
  - Archive upload & publication

**Custom workflow example**



~~~
### --- STEP 1: LOAD AND PREPARE DATA --- ###

```{r}
library(eDNAbridge)

data <- wl_get_all_wilderlab_data(public_only = FALSE) |>
  wl_map_wilderlab_data() |>
  wl_inject_dwc_wilderlab_constants()

head(data)
```

### --- STEP 2: VALIDATE AND PREPARE --- ###

```{r}
val_validate_dwc_to_level(data, "required")
val_validate_no_additional_terms(data)

issues <- data |>
  val_validate_all_fields()

issues |> dplyr::filter(error_level == "error")

dwc_frames <- data |>
  gen_tibble_to_dwc_frames()

head(issues)
```

### --- STEP 3: GENERATE ARCHIVE --- ###

#### -- STEP 3a: Generate EML file -- ####

```{r}
eml_base() |>
  eml_set_title("Example Dataset")
~~~

# Data Uploaded in the *eDNAbridge* pilot project

| **Organisation** | **eDNA samples** | **Occurrences** |
|---|---:|---:|
| DOC | 2,710 | 430,886 |
| HBRC | 1,753 | 310,699 |
| Earth Sciences | 63 | 22,548 |
| **Total** | **4,526** | **764,133** |

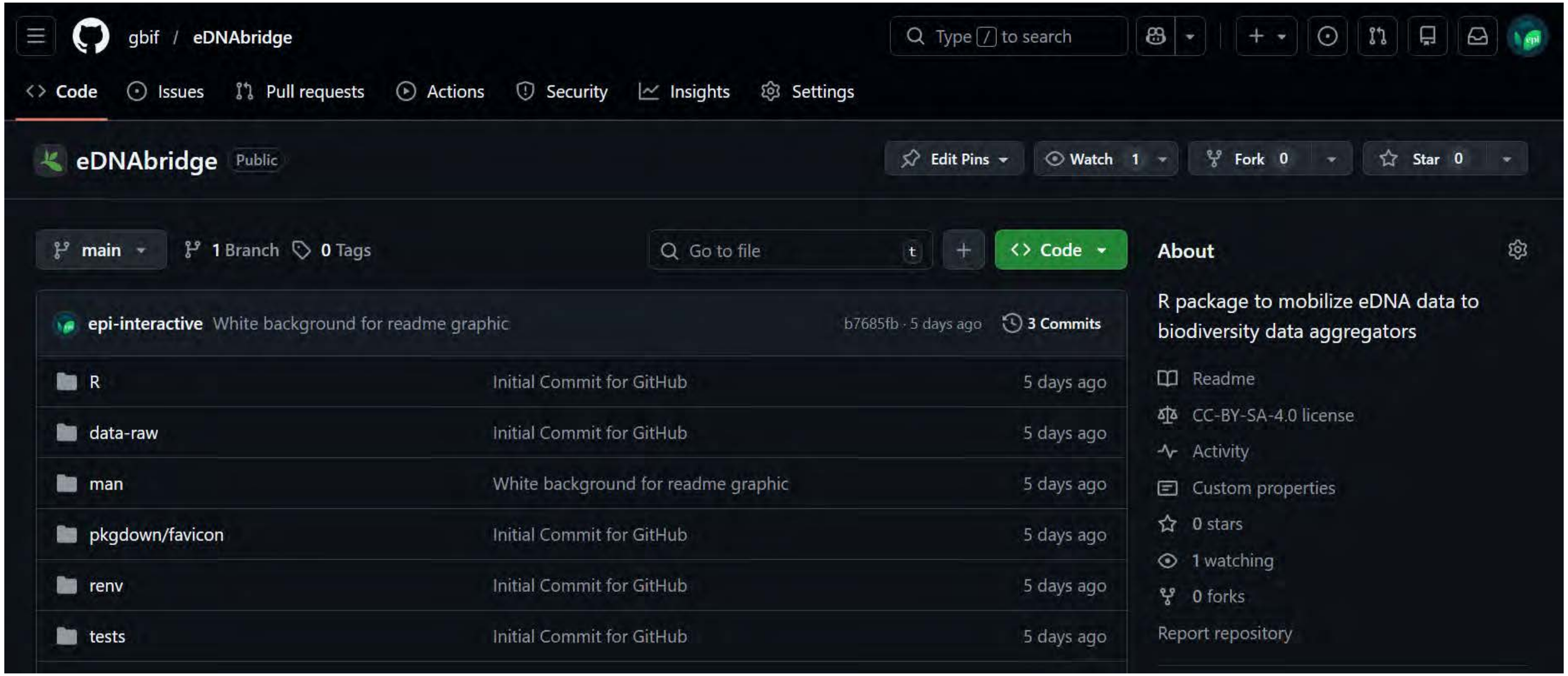
Core package tailored for all types of occurrence data, not just eDNA

Modular design allows new extensions in the future:

- FAIRe ingestion / validation?
- GBIF MDT server support?
- OBIS server support?

# Package availability

- Via GBIF’s GitHub – https://github.com/gbif/eDNAbridge



**epi-interactive** White background for readme graphic b7685fb · 5 days ago 3 Commits

| | | |
|---|---|---|
| R | Initial Commit for GitHub | 5 days ago |
| data-raw | Initial Commit for GitHub | 5 days ago |
| man | White background for readme graphic | 5 days ago |
| pkgdown/favicon | Initial Commit for GitHub | 5 days ago |
| renv | Initial Commit for GitHub | 5 days ago |
| tests | Initial Commit for GitHub | 5 days ago |

R package to mobilize eDNA data to biodiversity data aggregators

# Acknowledgements

WILDERLAB

Ministry for the
Environment
Manatū Mō Te Taiao

GBIF

Department of
Conservation
*Te Papa Atawhai*

HAWKES BAY
REGIONAL COUNCIL
TE KAUNIHERA Ā-ROHE O TE MATAU-A-MĀUI

Earth Sciences
New Zealand

# Thank you!

Presenters: Fiona Hodge and Kiel Hards

Other similar examples on Epi website:

www.epi.group

epi

posit

Posit Professional Services Partner

Certified B Corporation