BioSDGTime: a new database facilitating inquiries on global biodiversity and sustainable development progress

Veronica F. Frans, Michigan State University

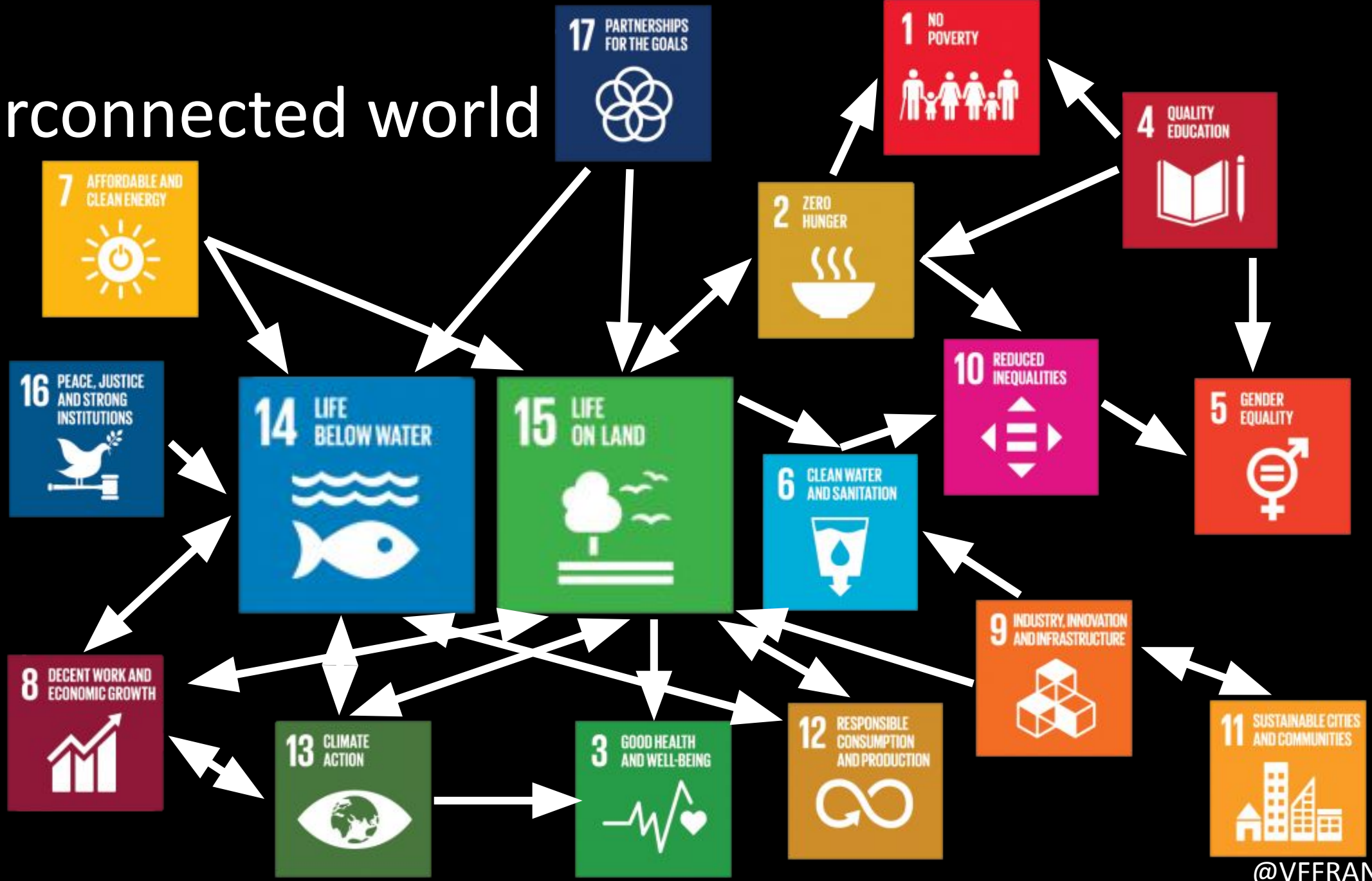
@VFFRANS

Saving our planet, people, and nature

Saving biodiversity

Interconnected world

Interconnected world



@VFFRANS

Data issue

Assessing the interconnected effects of sustainable development progress on biodiversity at global scales is a massive undertaking.

Data issue

We have biodiversity data, and sustainable development data, but they are not easily accessible—especially together.

Data solution

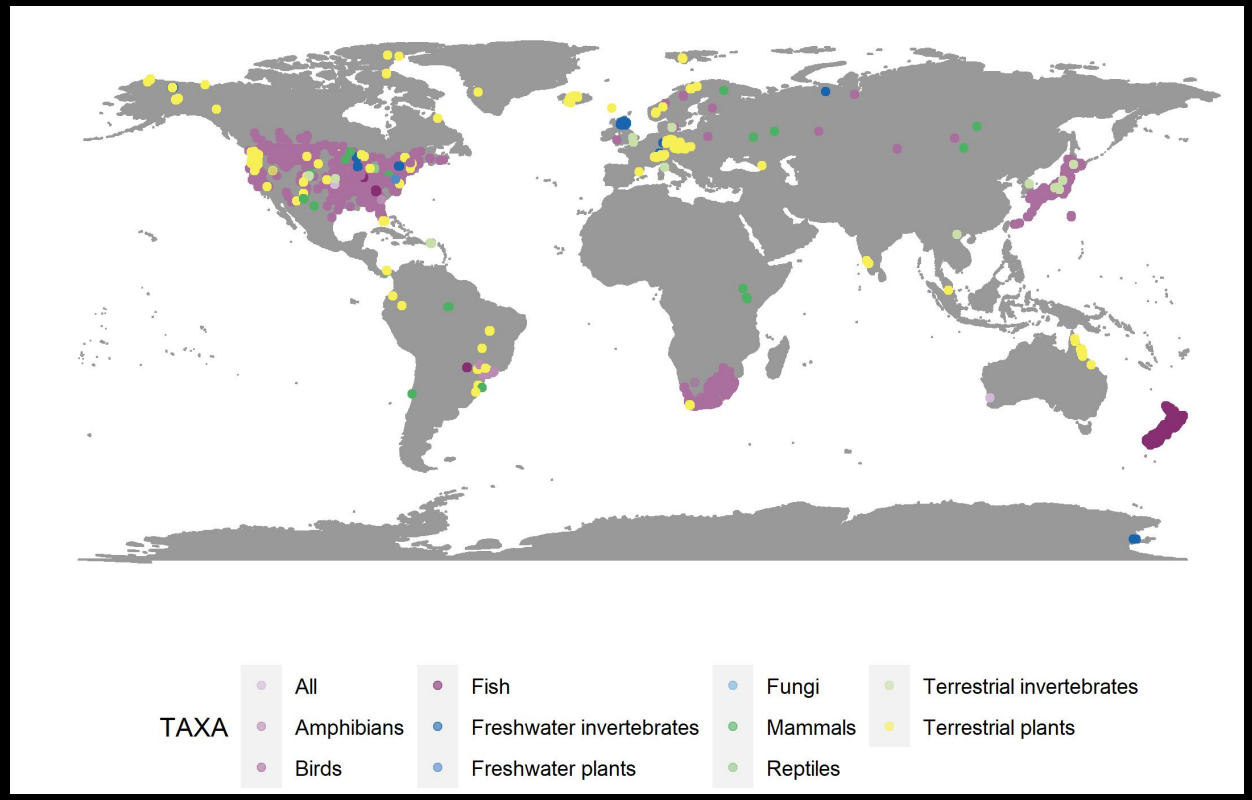
Cloud computing can integrate and update large-scale, complex datasets and make them easily available and accessible to researchers around the world in real time.

Aim

- To integrate SDG progress data with biodiversity data over time to create a BioSDGTime database.

Steps

- Discovery
- Planning



Discovery

The tools

Azure

R

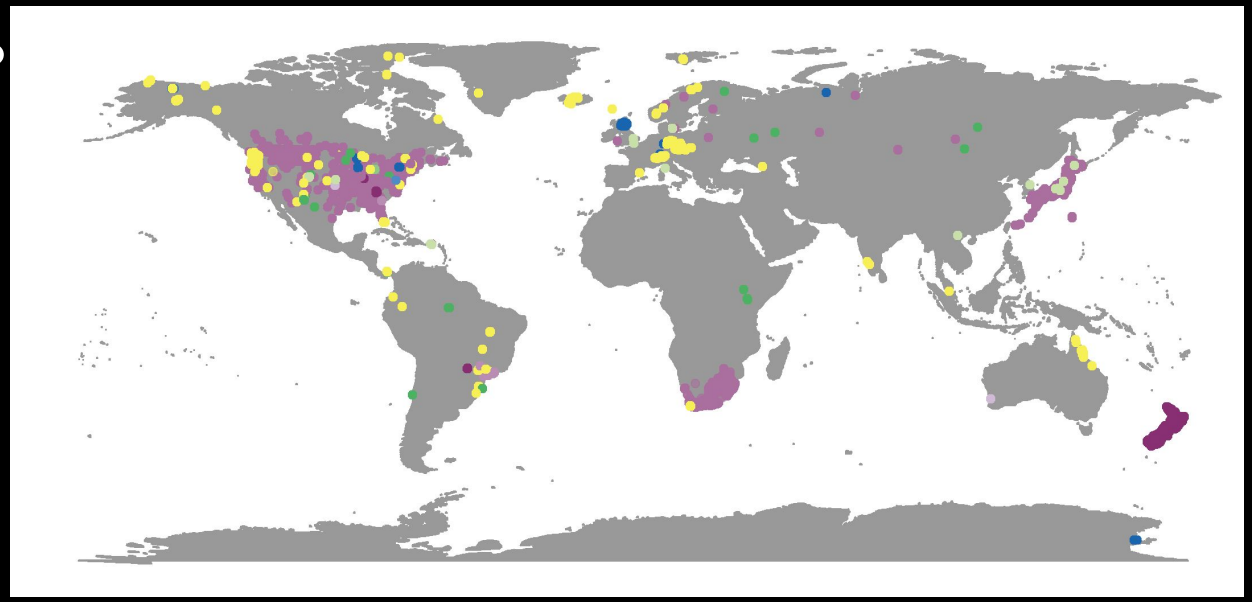
AzureR

PostgreSQL

The data

- UN SDGs
  - 232 SDG indicators for 260 countries/regions/areas
  - >1.5 million data points (>80 years)
  - UN, WorldBank (API's)
- Biodiversity
-

BioTIME



Discovery

The tools

Azure

R

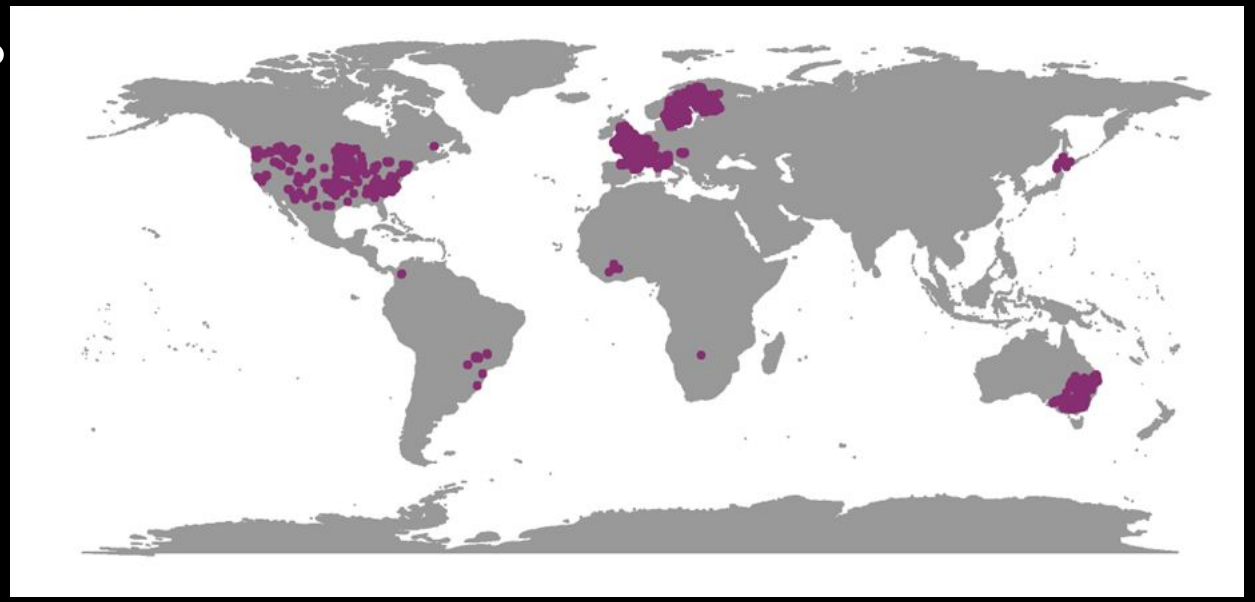
AzureR

PostgreSQL

The data

- UN SDGs
  - 232 SDG indicators for 260 countries/regions/areas
  - >1.5 million data points (>80 years)
  - UN, WorldBank (API's)
- Biodiversity
-

RivFishTIME



Discovery

The tools

Azure

R

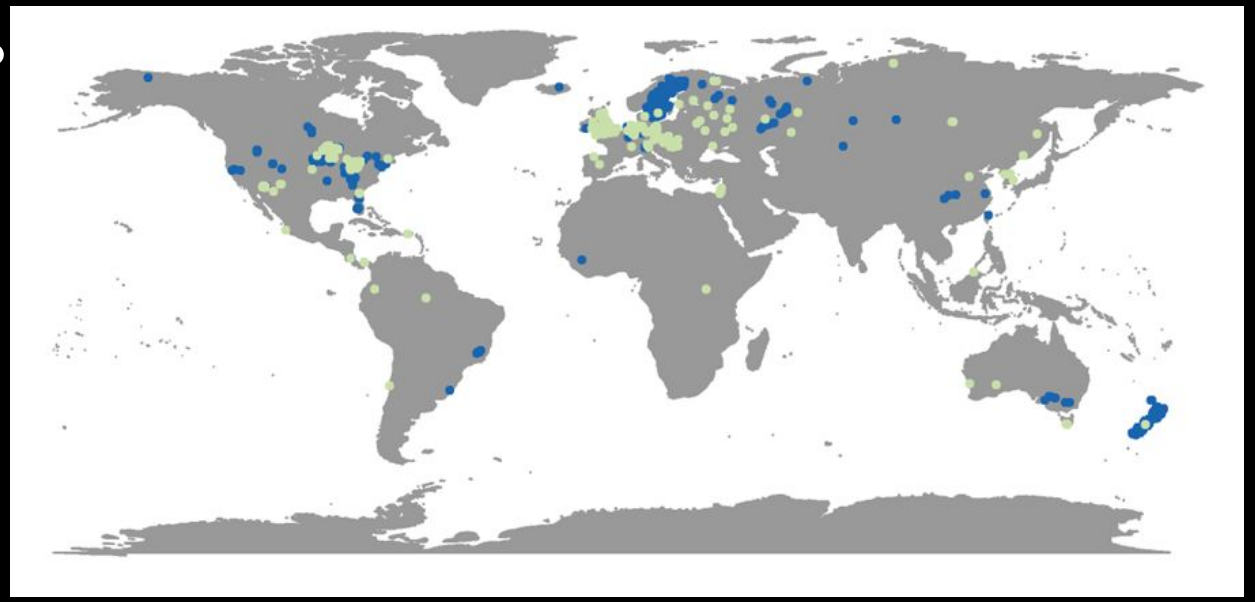
AzureR

PostgreSQL

The data

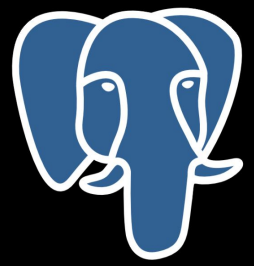
- UN SDGs
  - 232 SDG indicators for 260 countries/regions/areas
  - >1.5 million data points (>80 years)
  - UN, WorldBank (API's)
- Biodiversity
-

InsectChange



Discovery

The tools



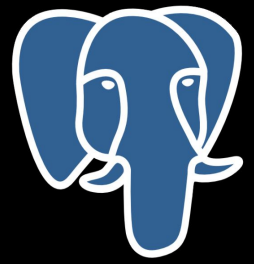
PostgreSQL

The reason

- SDG and biodiversity databases are constantly being updated.
- Relational databases store data in a structured way.
- As data is updated via APIs and on the cloud, the database relationships don’t change.
- Data can be accessed on-demand, not like a batch job on an HPCC.

Discovery

The tools



PostgreSQL

The reason

- Open science!

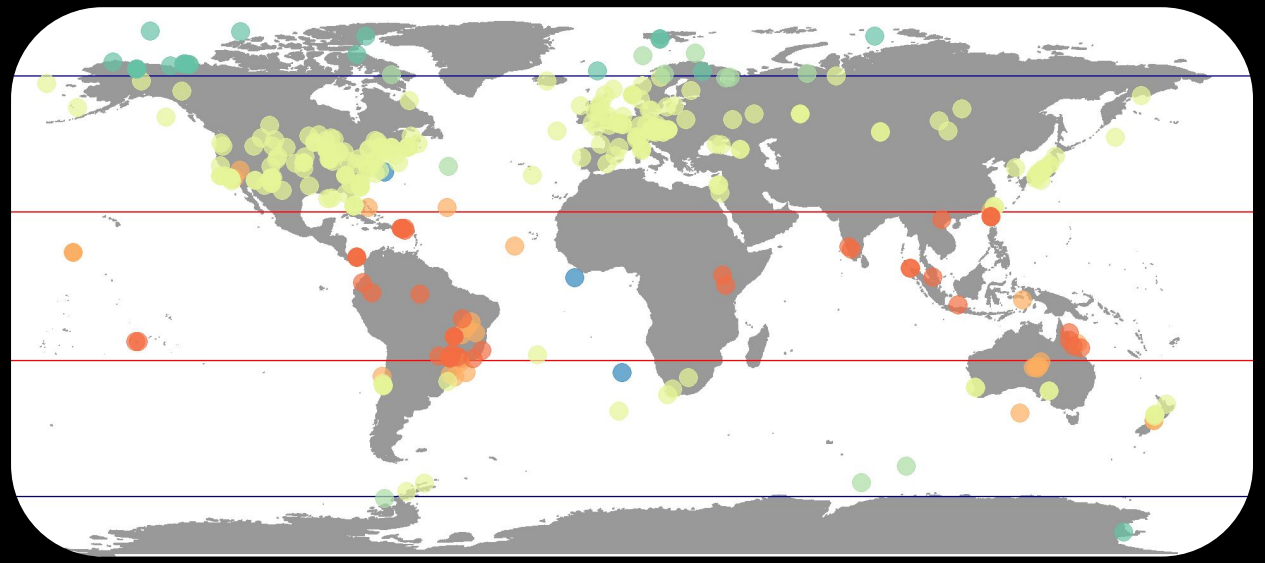
- Wrangling complex data, complex datasets – why do this once for a single study, if others can use it, too?
- With a good front-end interface, future researchers can access this dataset for modeling or even run models via the cloud.

Planning: Future goals

Data collection and onboarding (Azure VM)

Relational Database (Azure Database for PostgreSQL)

Set up and host access protocol



@VFFRANS

Discussion

- Challenges
  - Timing matters when you dream big.
- Successes
  - Cloud computing is a shift in thinking.
    - Research and results Impact Putting tools in others' hands.

Thanks!

@VFFRANS