

# Developing a collection data management service

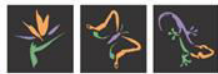
Brenda Daly

15 August 2018

Joint SANBI Biodiversity Information Management  
& Foundational Biodiversity Information Programme

**FORUM** 13–16 August 2018

**SANBI**  
Biodiversity for Life  
South African National Biodiversity Institute



science  
& technology  
Department:  
Science and Technology  
REPUBLIC OF SOUTH AFRICA



National  
Research  
Foundation

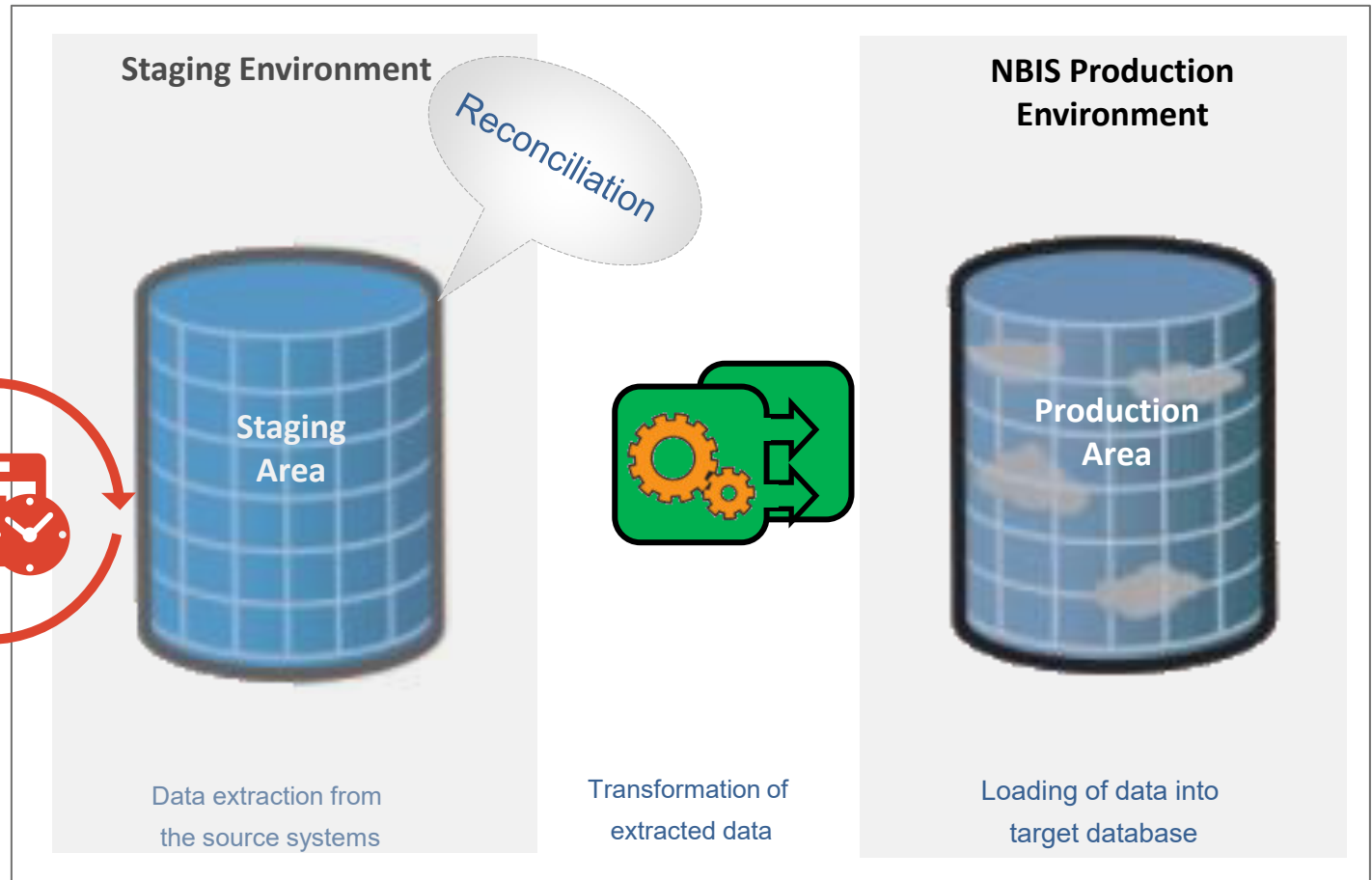
# Advantages of data aggregation

- More data is better – informed decisions
- Decision-making is faster if aggregated data already exists (BioGaps)
- Expose patterns and trends
- Option for making detailed analyses
- Investment in processes and infrastructure and provide solutions for long-term hosting

# By-products of data aggregation

- Allows researchers to question the results of the separate analyses based on independent lines of evidence
- Backup
- Institutions can have their own portal and share
- Data upgrade quality by comparison of duplicate specimens (e.g. determination differences)
- Potential exposure of errors (crowd-source data quality checks)
- Collective knowledge

# National Data Stores



# Error-correction flow



Data Providers supplying edits  
and new records  
asynchronously

## Quality Control:

- Species names
- Latitude and longitude
- Collecting dates
- Original and processed records for data loss

# Aggregators?

## Zoological

- Symbiota
- ALA
- Specify
- Grow your own

## Botanical

- BRAHMS

# Challenges

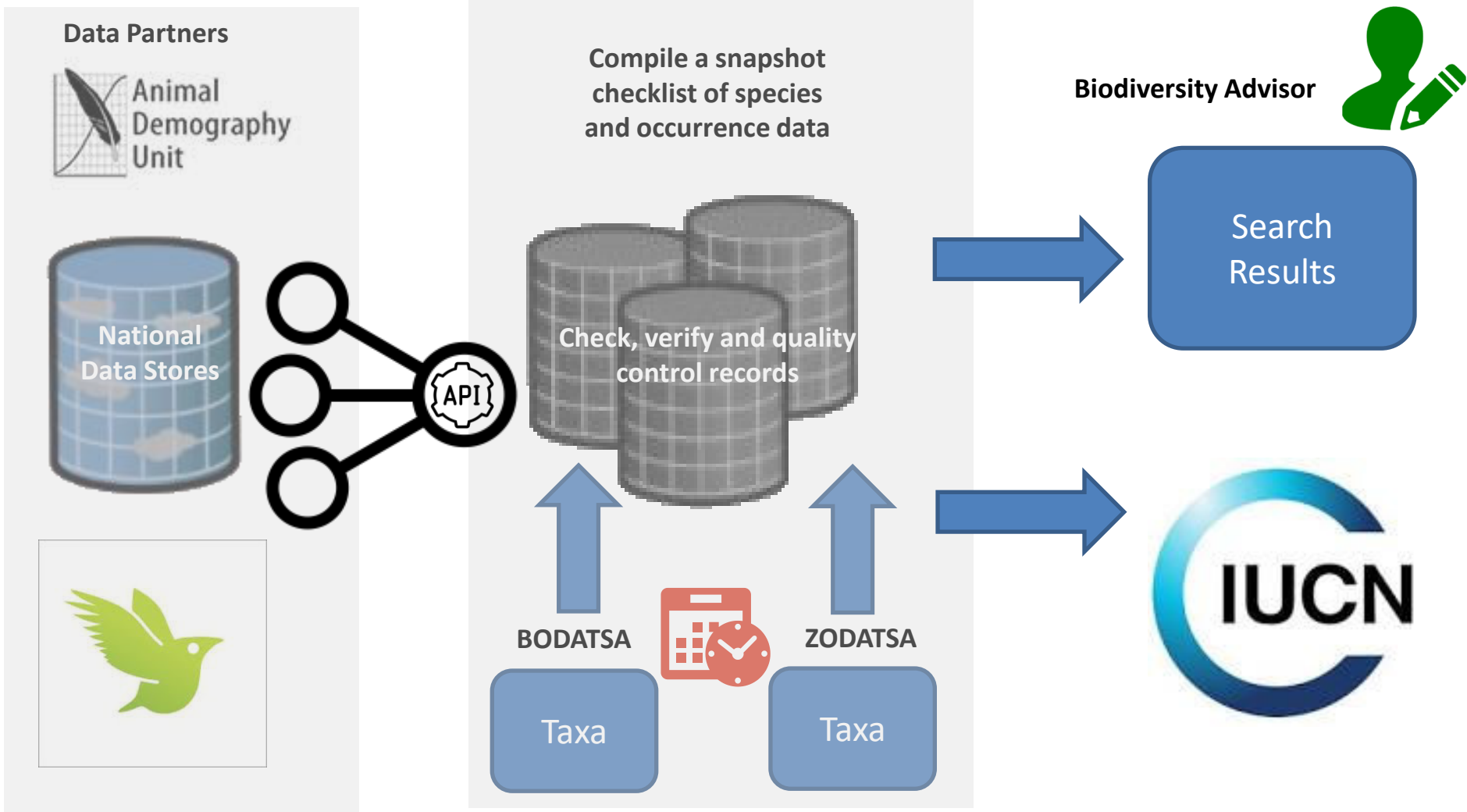
- Closed-source non-API-driven desktop applications
- Name changes and Type specimens
- 2 or 3 versions of the same record (to do: list fields)
- Taxonomic Services Backbone
- Persistent identifiers at a record-level and taxa

# What needs to be done?

- Integration and alignment:
  - Taxonomic Services Backbone
  - Persistent identifiers (collection and record identifiers)
  - Sensitive data
  - Metadata (Provenance)
  - Attribution
  - Gazetteers
- Data checking (query data providers when errors are detected)
- Rigorous standards (NSCF)
- Identify duplicate records and inconsistencies



# Red List Assessment System



**SANBI**

Biodiversity for Life



?

1

Thank you