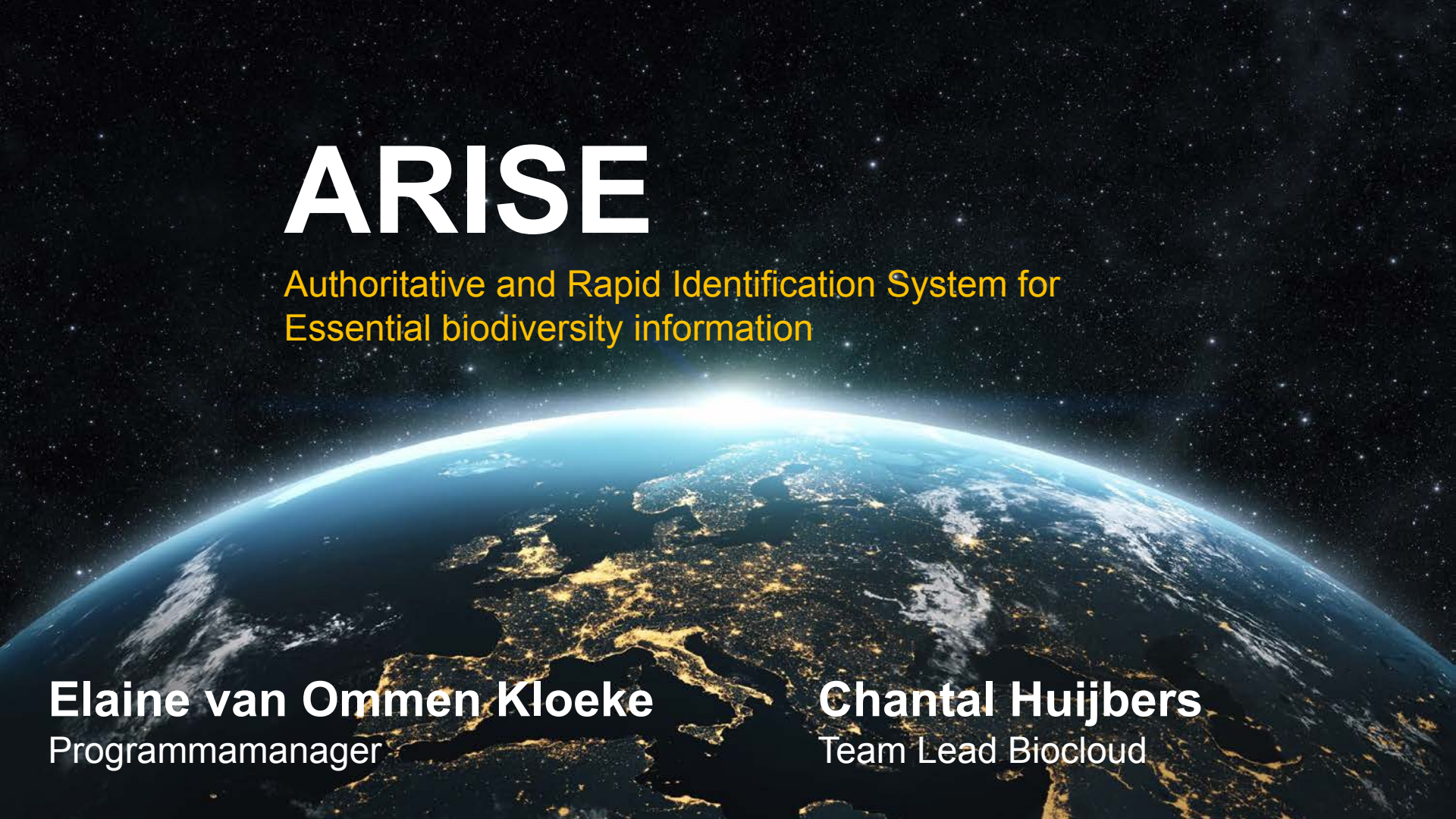


ARISE

Authoritative and Rapid Identification System for Essential biodiversity information

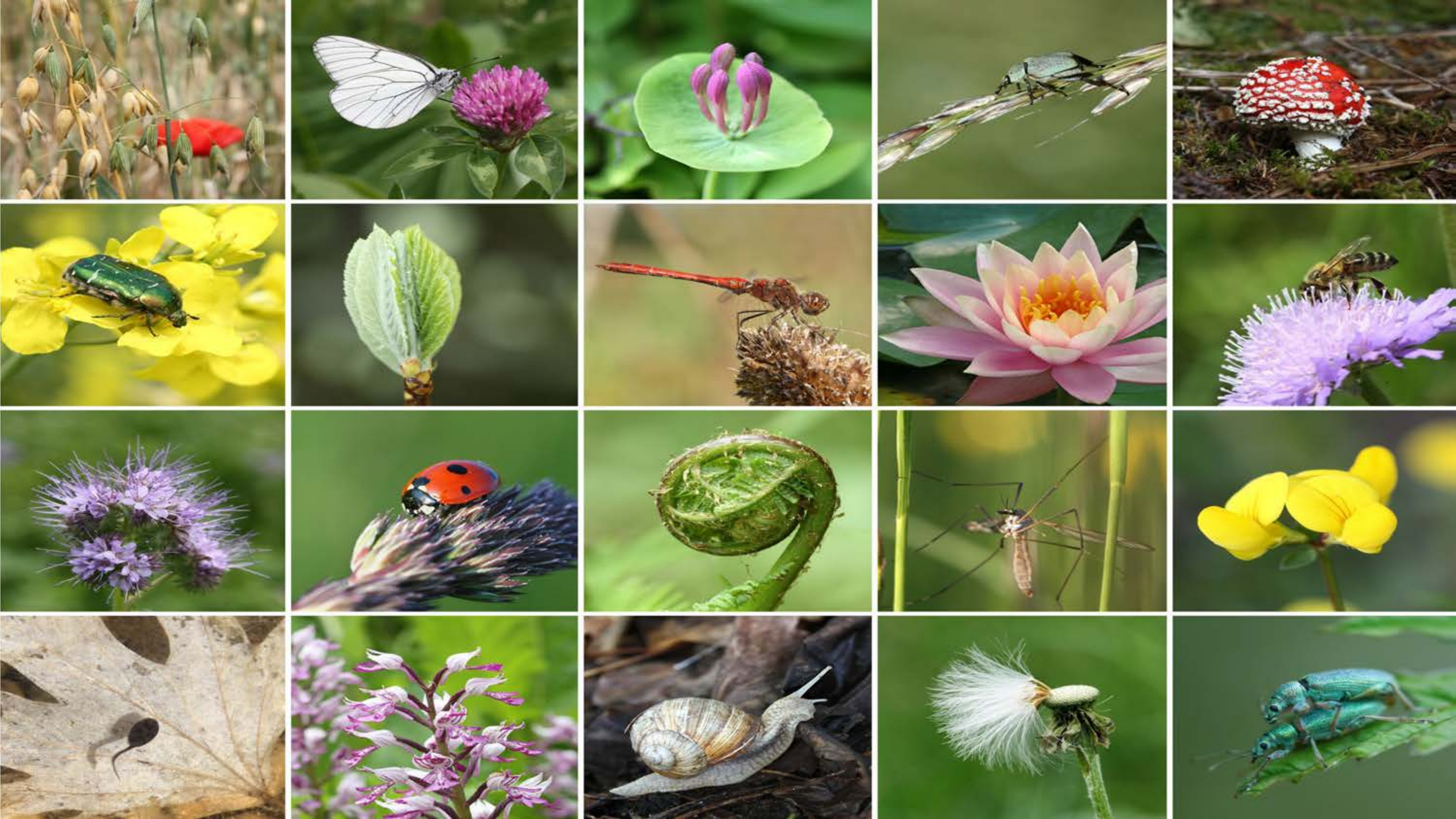
Elaine van Ommen Kloeke
Programmamanager

Chantal Huijbers
Team Lead Biocloud

UNIVERSITY
OF AMSTERDAM

WESTERDIJK
FUNGALBIO
DIVERSITY
INSTITUTE

UNIVERSITEIT TWENTE.



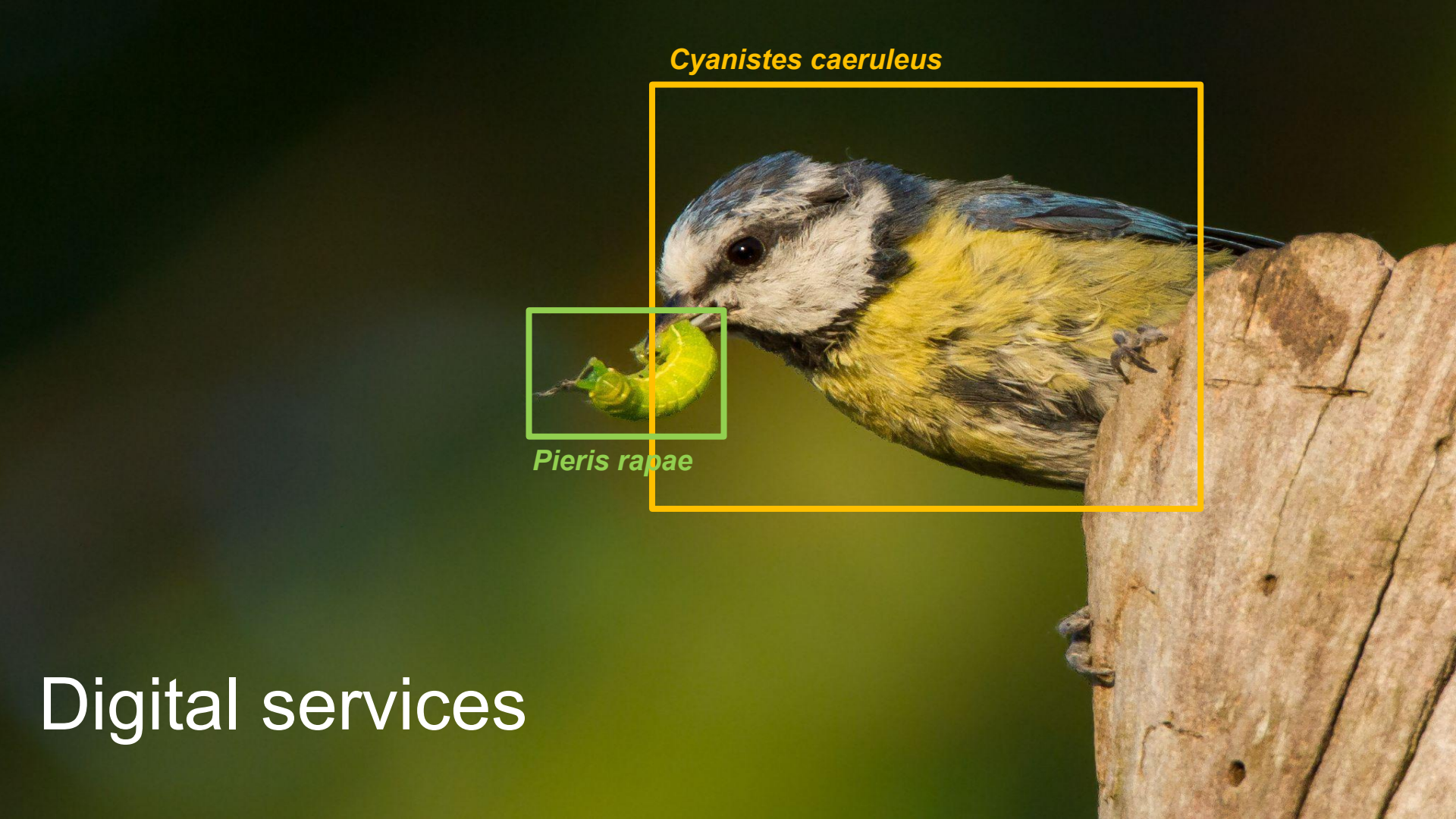

Food


Climate


Underground

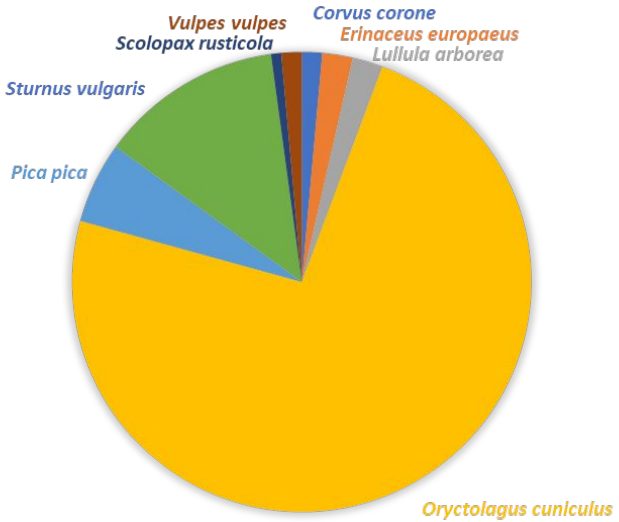
Accessible

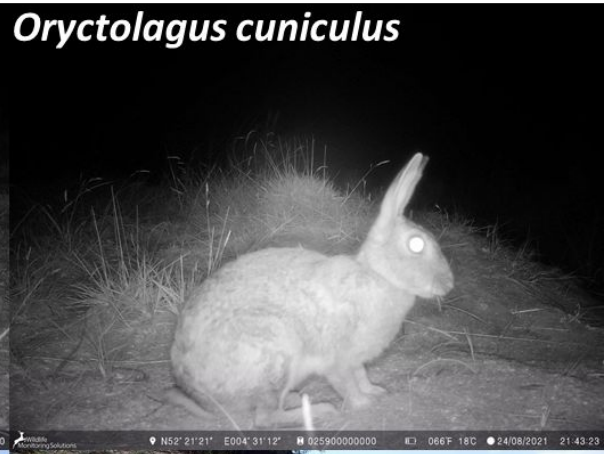

Innovate



Digital services

Amsterdamse Waterleiding Duinen






Wildlife Monitoring Solutions
N53° 09′ 04″ E006° 29′ 40″
210800000000
074°F 23°C
21/03/2022 15:51:35


Connected

Interoperability

Biodiversity

About us Tools

Search, Analyze, Discover

A new way of discovering species.

Discover Start Searching

Search Advanced search

Search...

Popular now: Name species, Grasshopper, Beetle

Our Tools

Overview of our tools

Our Tools

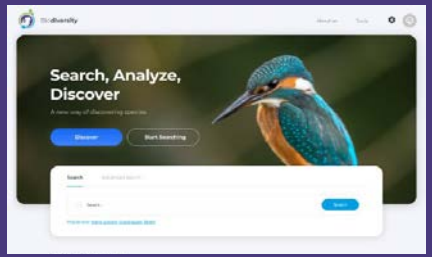
Access to data and services

Waarneming.nl
DISSCO
BOLD SYSTEMS
GBIF
DIOPSIS
Catalogue of Life
xeno-canto
LifeWatch ERIC
AGOUTI
SURF

Linked to other infrastructures

Linked data
M2M connections

ARISE platform
'a one-stop-shop for all your biodiversity data services'



Back end facilities

- Data management & storage
- Access control
- API services

End user services

- Explore / Search data
- Import / Link data
- Annotate data
- Explore algorithms
- Construct and train algorithms
- Run algorithm to identify species
- View algorithm results

The Future of ARISE

Building & connecting to other infra's

Programme organization

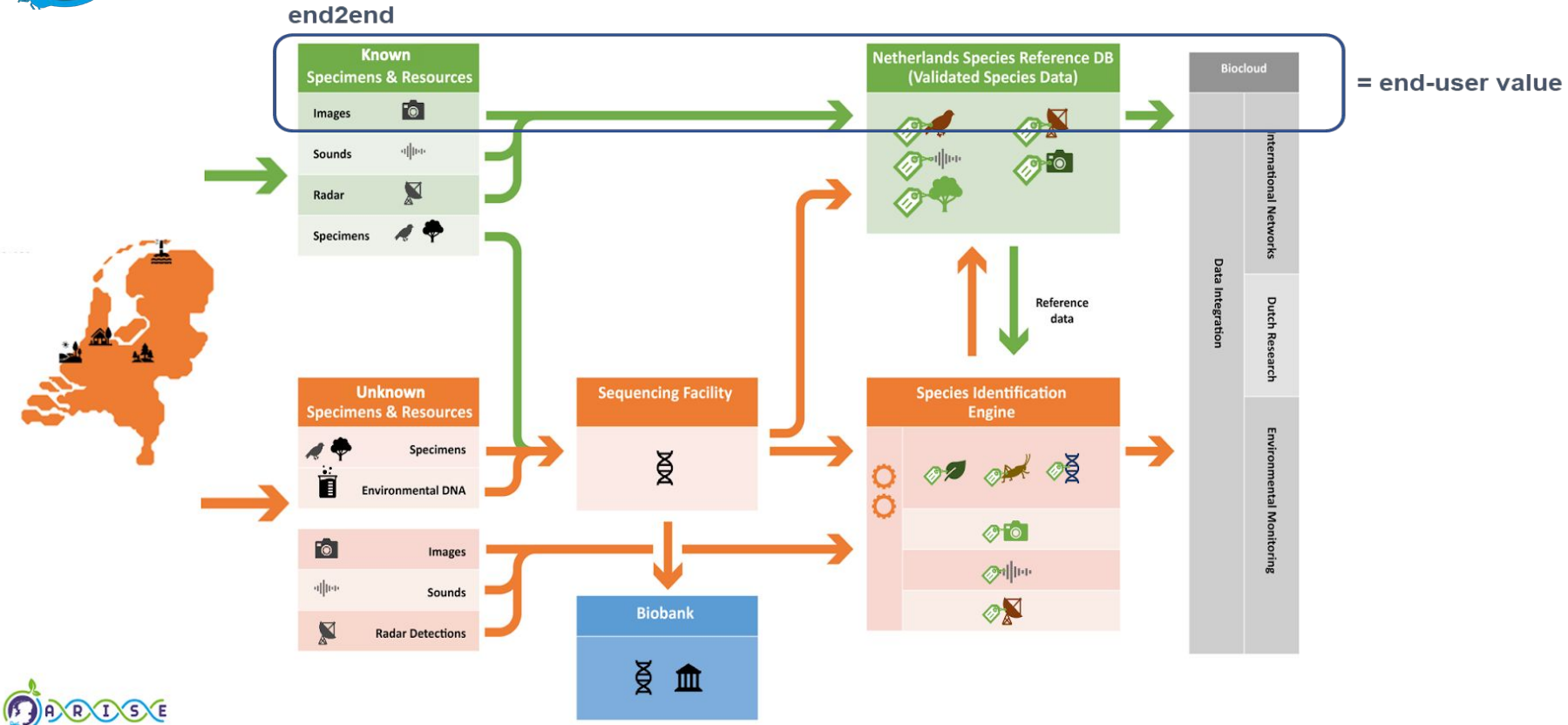
Multidisciplinary & cross-partner

5 dedicated multidisciplinary & multi-organization teams

Build on existing knowledge, data and services

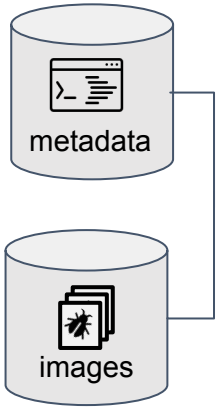
User stories to drive focus, deliver end2end, and iterative development

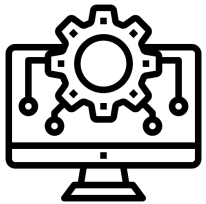
ARISE architecture design

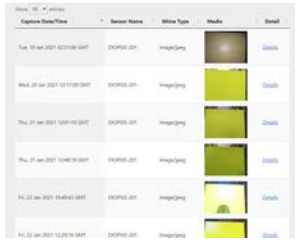


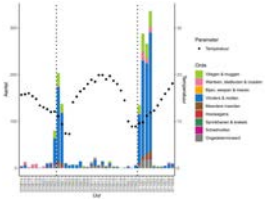
Use case: Diopsis insect cameras

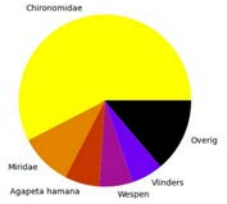
Collect → Store → ID → Expose → Explore





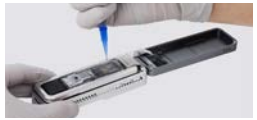


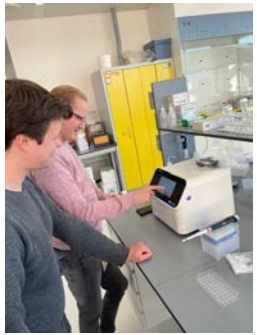


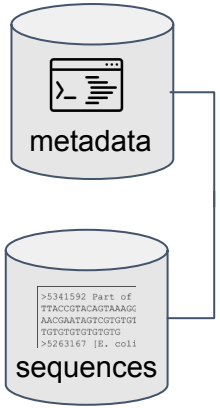


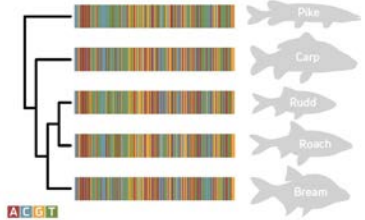
Use case: species barcoding

Collect → ID → Store → Expose → Explore



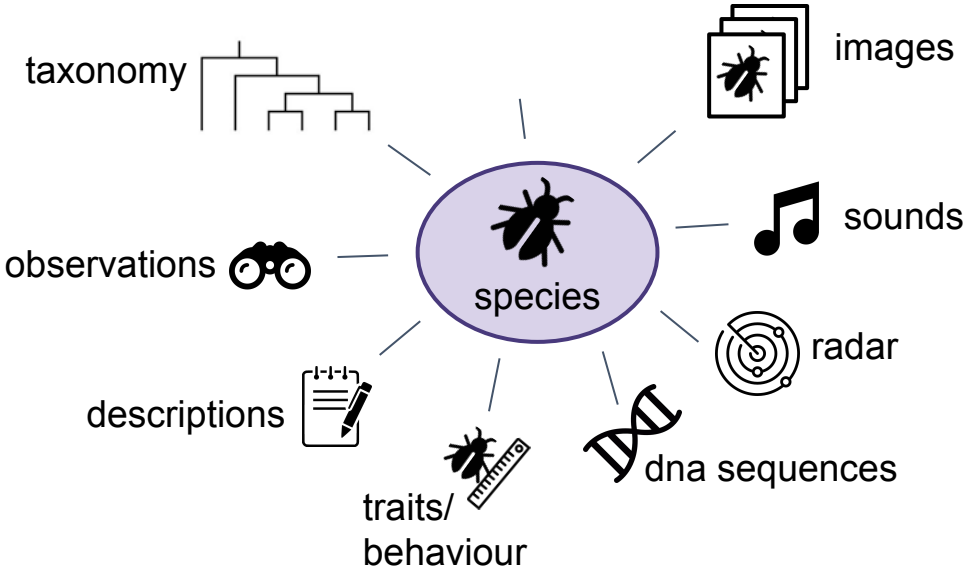






Biocloud

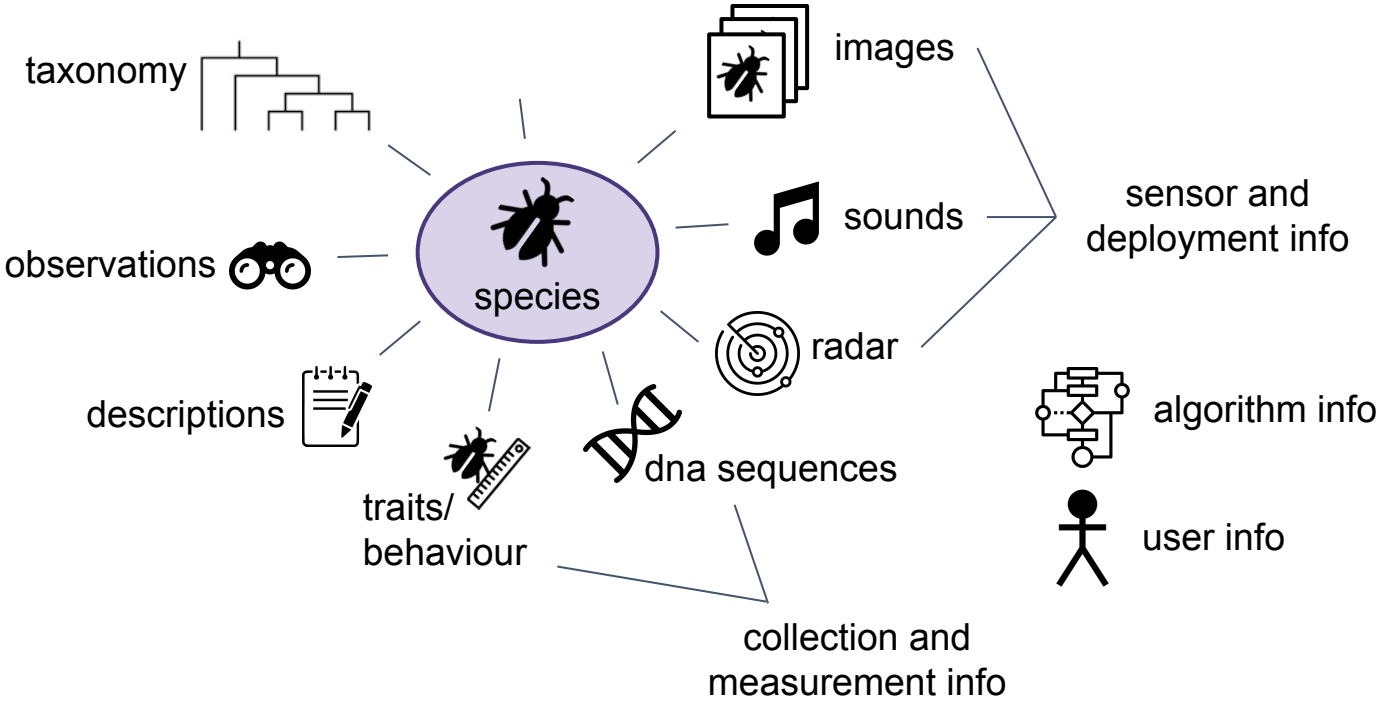
A scalable data management system for petabytes of data of various kinds, with linked metadata



Species Reference Database

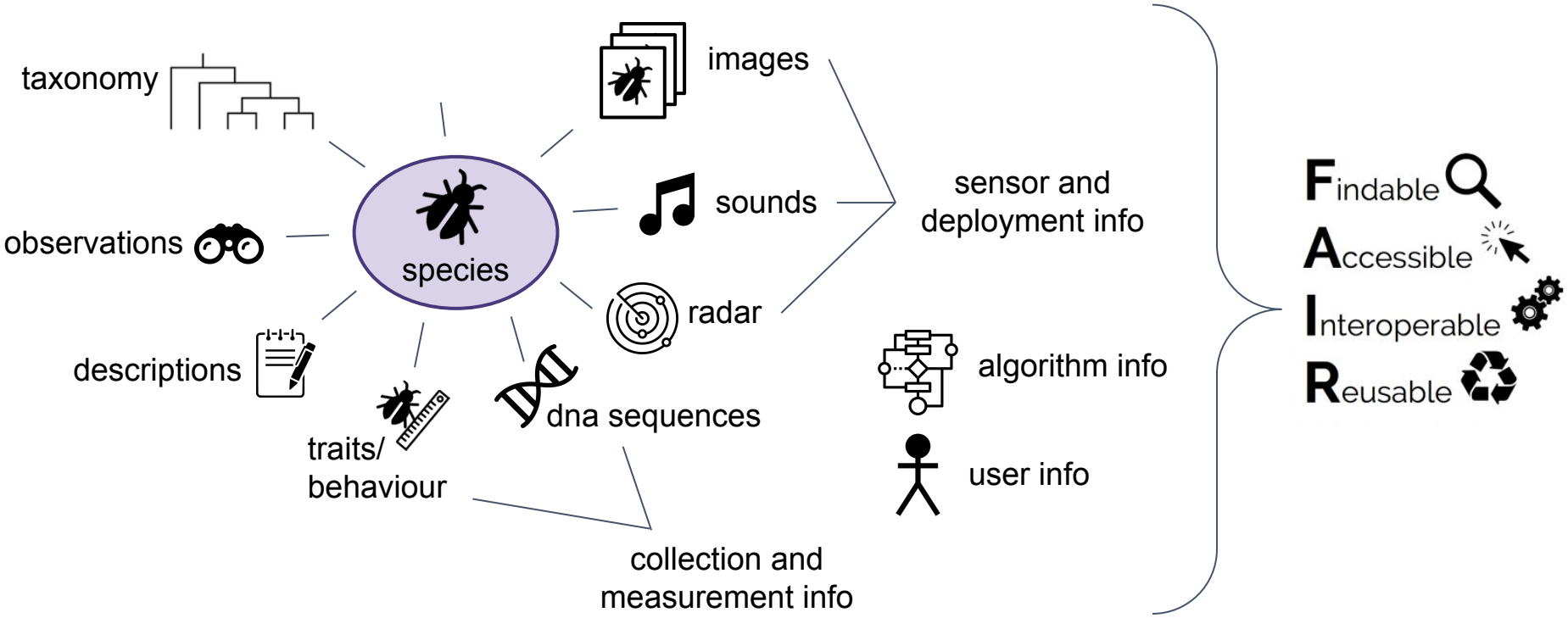
Biocloud

A scalable data management system for petabytes of data of various kinds, with linked metadata



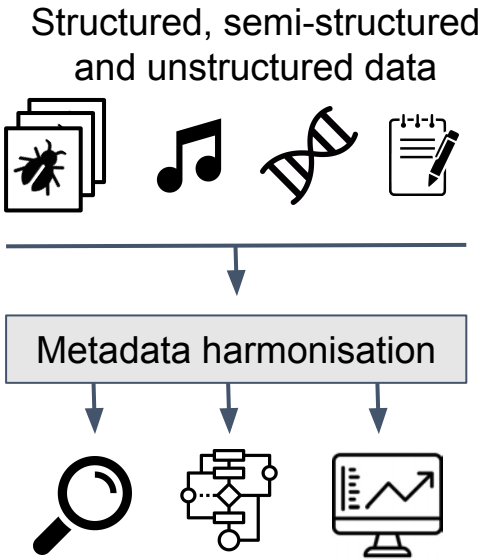
Biocloud

A scalable data management system for petabytes of data of various kinds, with linked metadata



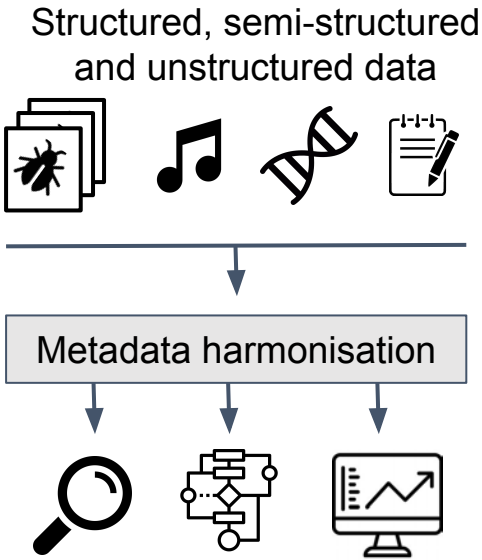
Biocloud

Work in progress:
architectural design and development of data management system



Biocloud

Work in progress:
architectural design and development of data management system

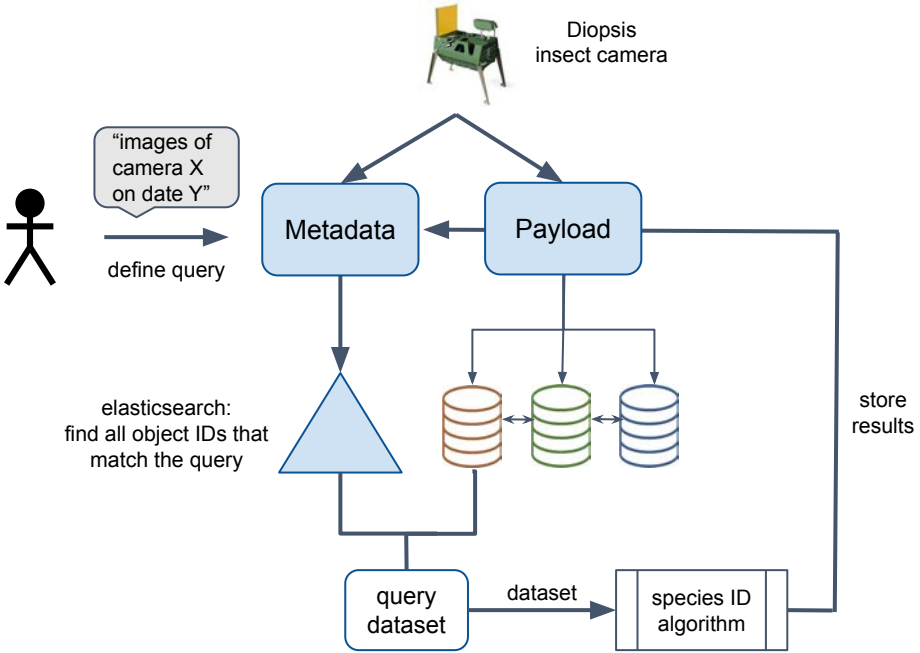


Minimal system requirements:

- Provenance and traceability
- Data objects must be uniquely identifiable by permanent identifiers
- External data objects reside in their original systems and are connected through identifiers and metadata
- Proper handling of ownership, access control and licensing

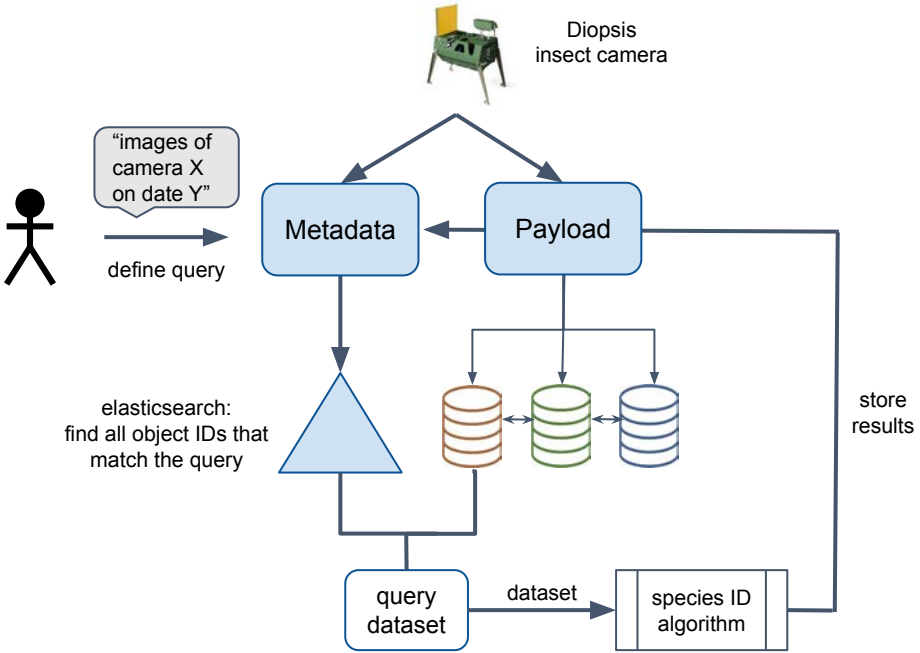
Biocloud

Work in progress:
architectural design and development of data management system



Biocloud

Work in progress:
architectural design and development of data management system

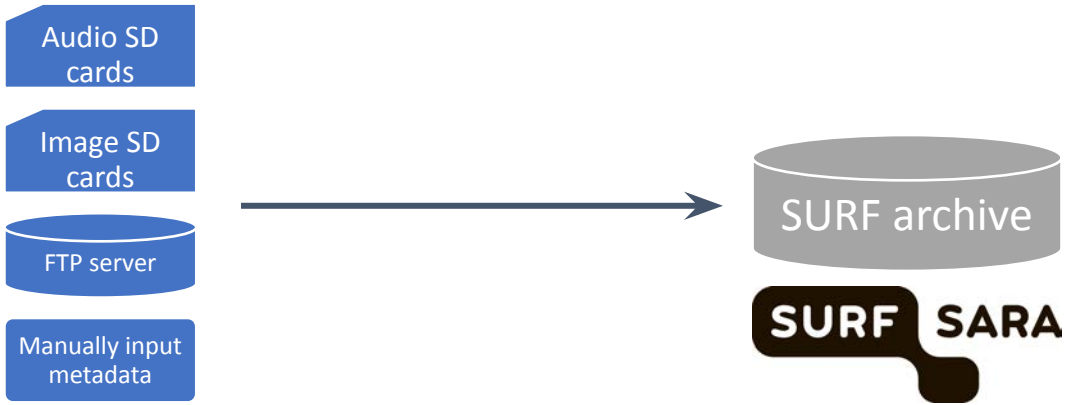


Data storage

A researcher's perspective

Objective: Get data collected at ARISE Monitoring Demonstration Sites into SURF data archive

- Collected by multiple people, from multiple sensors, in multiple formats
- Large volumes, ~15TB from 48 wildlife cameras & 26 AudioMoth over 7 months

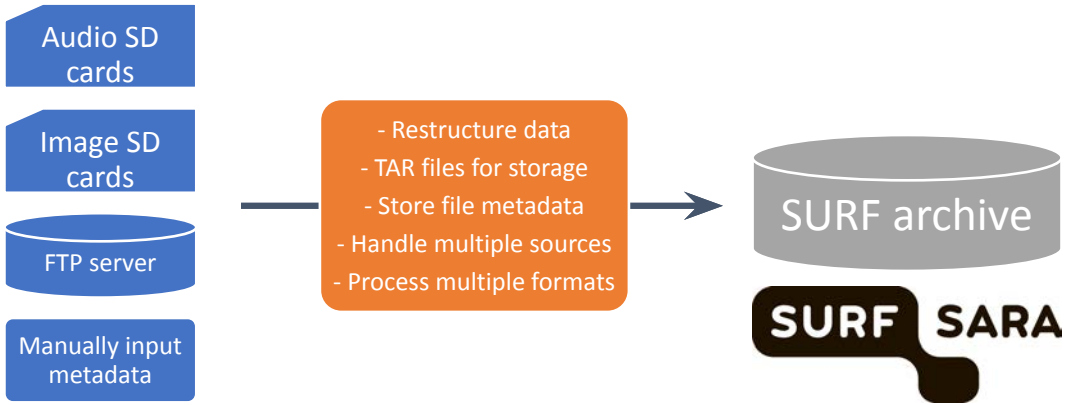


Data storage

A researcher's perspective

Before archiving data must be:

- Restructured to standard naming conventions, folder structure
- Placed in Tape archive (TAR) format (>1 gb)
- Have appropriate metadata appended

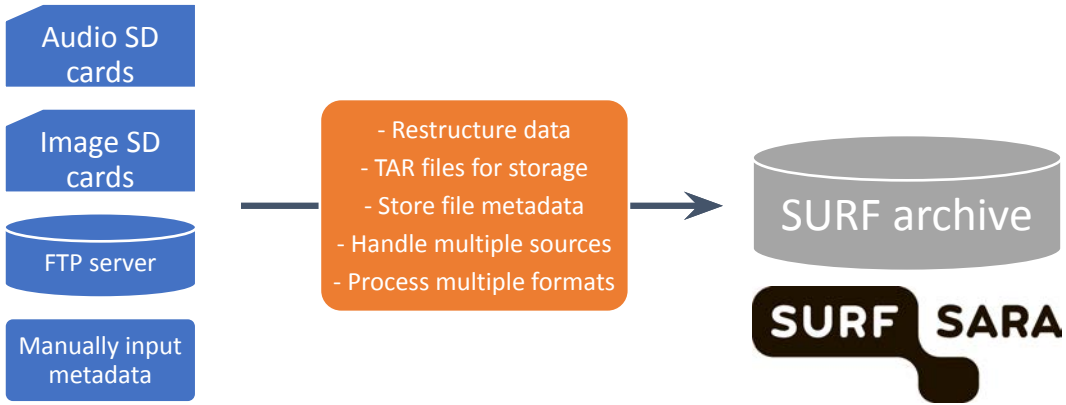


Data storage

A researcher's perspective

These processes needed to be:

- As automatic as possible
- Usable by multiple people
- User friendly

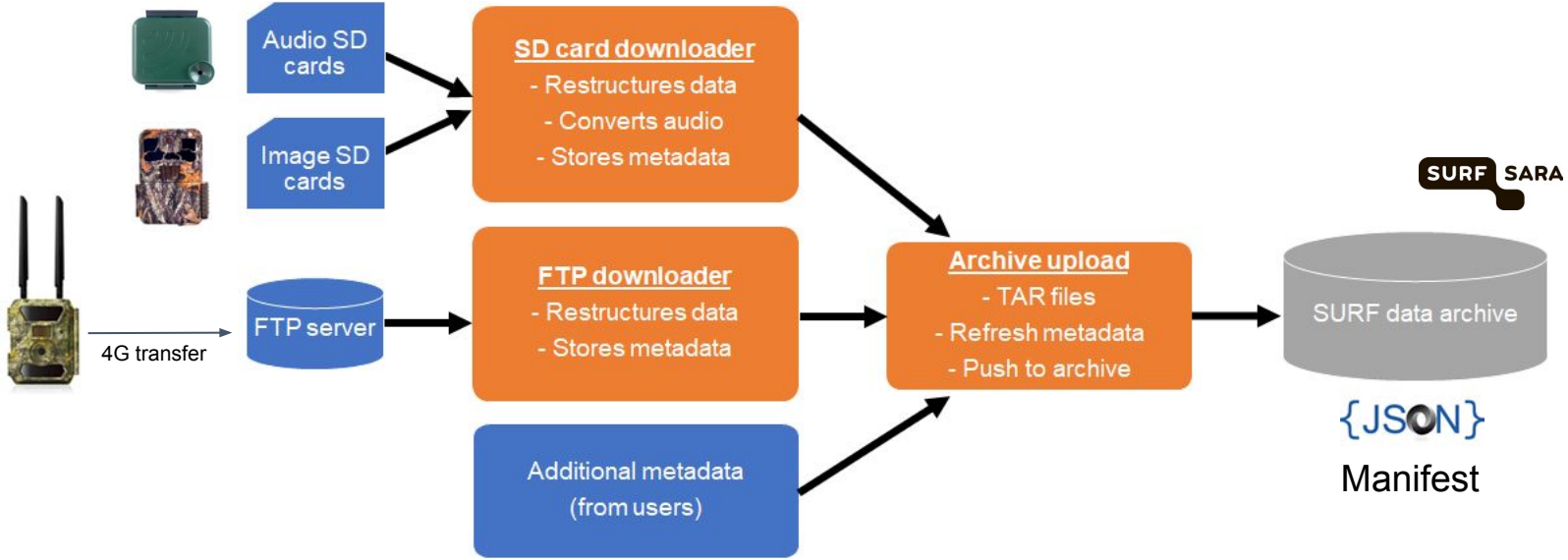


Data storage

A researcher's perspective

Current solution developed by ARISE Data Scientist at UvA

Digital sensor → **Short-term storage** → **Download** → **TAR & Upload** → **Long-term storage**



Data storage

A researcher's perspective

Custom software to:

- allow bulk download of data
- restructure data from the different sources
- TARring of data

  > Each TAR file contains JSON metadata, manifest file that details content, folder structure with files
- upload to archive
- creation and updating of manifests for different data types

While this all sounds relatively simple, building systems to do all this is not trivial for many users!

Possible cloud solutions?

What are your thoughts on these challenges?

Data Storage

- As an ecological researcher, how do I find the right storage for my large volume of data?
- How to keep track of what is uploaded in TARs?
- How to make this more user friendly?

Linking data between systems

- How to ensure interoperability between data stored and managed in different systems?
- How to tackle data quality issues, such as inconsistency, inaccuracy, out-of-dateness, incompleteness?
- Data lakehouse solutions: how to decide which technology is best?

Data to compute, or compute to data?

- How to bring data close to computational services?
- When is it better to bring the compute to data?

Stay up to date!

Subscribe to our newsletter:

www.arise-biodiversity.nl/using-arise

elaine.vanommenkloeke@naturalis.nl

UNIVERSITEIT TWENTE. Naturalis Biodiversity Center WESTERDIJK FUNGALBIO DIVERSITY INSTITUTE UNIVERSITEIT VAN AMSTERDAM