UNIVERSITY OF AMSTERDAM
Faculty of Science

LifeWatch ERIC

IBED
Institute for Biodiversity and Ecosystem Dynamics

Biodiversity data workflows for Digital Twins

W. Daniel Kissling[1,2]

[1]Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam

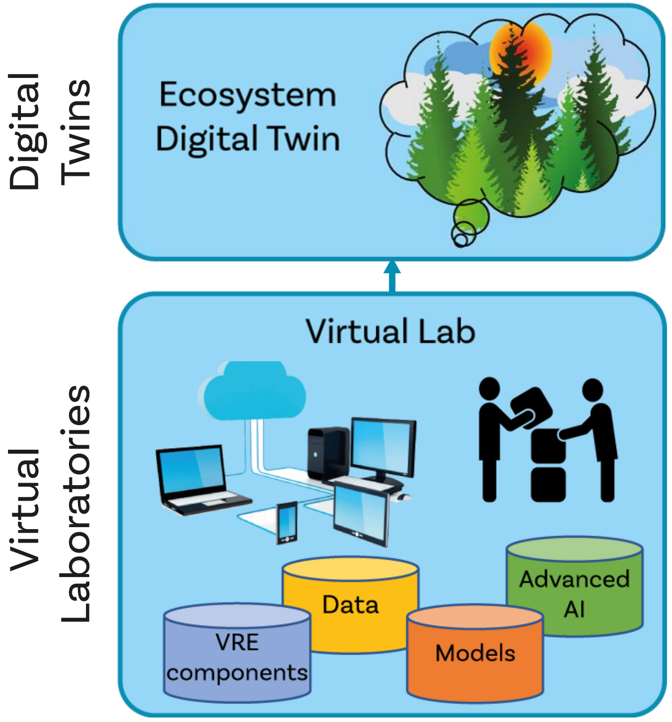
[2]Virtual Laboratory & Innovations Center (LW-VLIC), LifeWatch-ERIC

Digital Twin Conference, Wageningen, 14 December 2022

Digital Twins of ecosystems

Simulation platform for complex (eco)system analysis →

Interactive, collaborative, virtual research environment →



- Digital replicas of entire ecosystems
- Supported by a cloud-based platform for digital modelling and simulation
- Allows scientists to predict interactions & responses to external drivers

AI and machine learning approaches to enable new theory discovery and novel insights from high-dimensional sparse data spaces

Process-based and Big Data-driven models

Infrastructure services from a Virtual Research Environment (VRE) (e.g. for data discovery, workflow management, cloud computing etc.)

Findable, Accessible, Interoperable, and Reusable (FAIR) data

Biodiversity data

Workflows

1 Primary observations

Opportunistic observations
- Incidental occurrence and abundance data from research and citizen science

Structured surveys
- Inventories & checklists from research and citizen science
- Long-term species monitoring

Trait sampling
- Morphological measurements
- Physiological measurements
- Phenology records

DNA sampling
- Intraspecific genetic diversity
- eDNA

Airborne remote sensing
- Imaging spectroscopy
- LiDAR

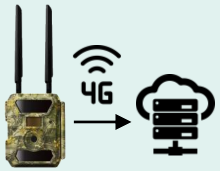
Satellite remote sensing
- Multispectral - thermal
- Radar
- Hyperspectral
- LiDAR

2 Data generation

Data integration and harmonization

- Data mobilization
- Data standardization
- (Meta)data standards
- Ontologies and vocabularies

Automated data streams



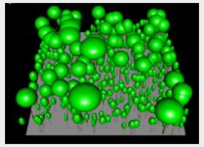
- Autonomous sensors
- Data transmission
- Data archiving & big data storage
- Remote sensor monitoring

Big data processing

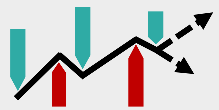
- Efficient, scalable and distributed computing
- HPC and cloud services
- Free and open-source software

3 Models and simulations

Simulations and trend analysis



Ecosystem simulation



Biodiversity change indicators

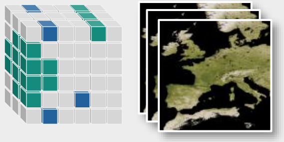
AI models

Detection, tracking, classification, segmentation

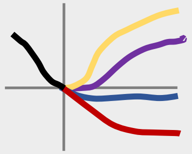
Scenarios and forecasts

Predictive models

$Y \sim f(x_1,\ldots,x_n)$

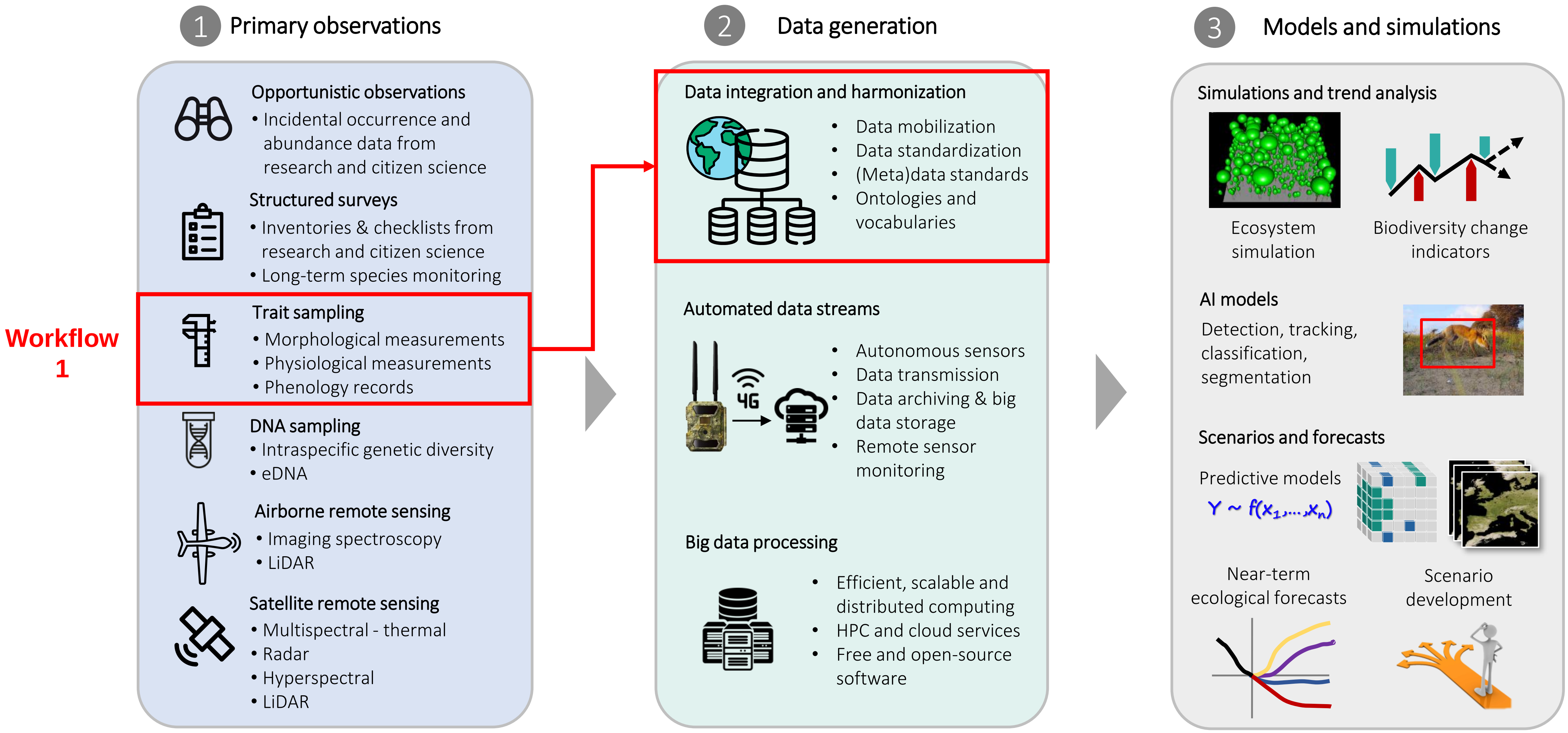


Near-term ecological forecasts



Scenario development

Biodiversity data



Workflow 1: Data integration & harmonization

Ecological Informatics 64 (2021) 101372

Contents lists available at ScienceDirect

ELSEVIER

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Check for updates

Integrating long-tail data: How far are we?

Kristin Vanderbilt [a,*], Corinna Gries [b]

[a] Environmental Data Initiative, University of New Mexico, Albuquerque, NM 87131, United States of America
[b] Environmental Data Initiative, University of Wisconsin, Madison, Madison, WI 53706, United States of America

Ecological Informatics 62 (2021) 101206

Contents lists available at ScienceDirect

ELSEVIER

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Check for updates

Integration and harmonization of trait data from plant individuals across heterogeneous sources

Tim P. Lenters [a], Andrew Henderson [b], Caroline M. Dracxler [a], Guilherme A. Elias [c], Suzanne Mogue Kamga [d], Thomas L.P. Couvreur [e], W. Daniel Kissling [a,*]

[a] Department of Theoretical and Computational Ecology (TCE), Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, P.O. Box 94240, 1090 GE Amsterdam, the Netherlands
[b] The New York Botanical Garden, Bronx, NY 10458-5126, USA
[c] Programa de Pós-Graduação em Ciências Ambientais, Universidade do Extremo Sul Catarinense UNESC, Av. Universitária, 1105, 88806-000 Criciúma, SC, Brazil
[d] Université de Yaoundé I, Ecole Normale Supérieure, Département des Sciences Biologiques, Laboratoire de Botanique systématique et d'Ecologie, B.P. 047 Yaoundé, Cameroon
[e] IRD, DIADE, Univ Montpellier, Montpellier, France

Long-tail data:

- small, tabular, site-based datasets
- collected by individual researchers
- often different in formats, units, semantics etc.
- little data curation and mostly not shared
- but contain very valuable information

Trait data integration workflow:

- harmonizes quantitative individual-level plant trait data
- captures data from heterogeneous sources (taxonomic revisions and ecological datasets)
- standardizes data using terms from existing ontologies

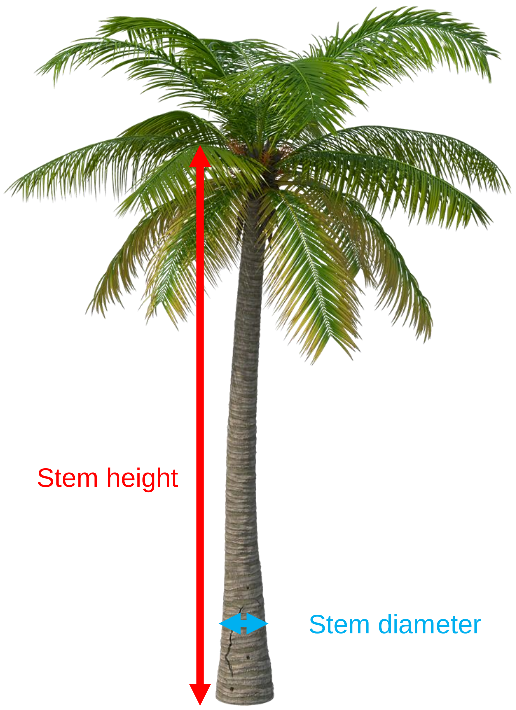
Workflow 1: Data integration & harmonization

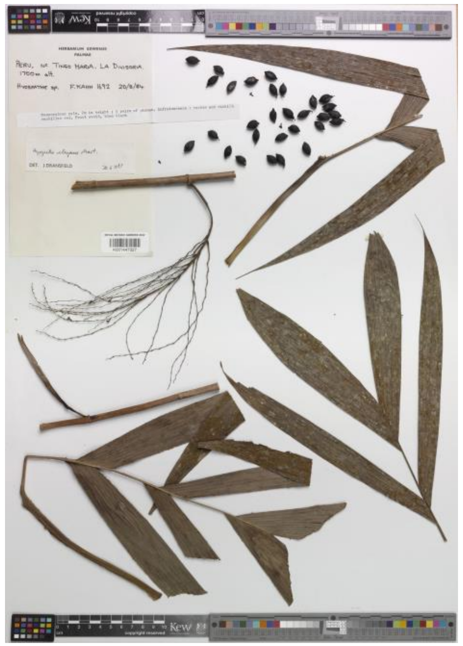
Spreadsheets from taxonomists and ecologists

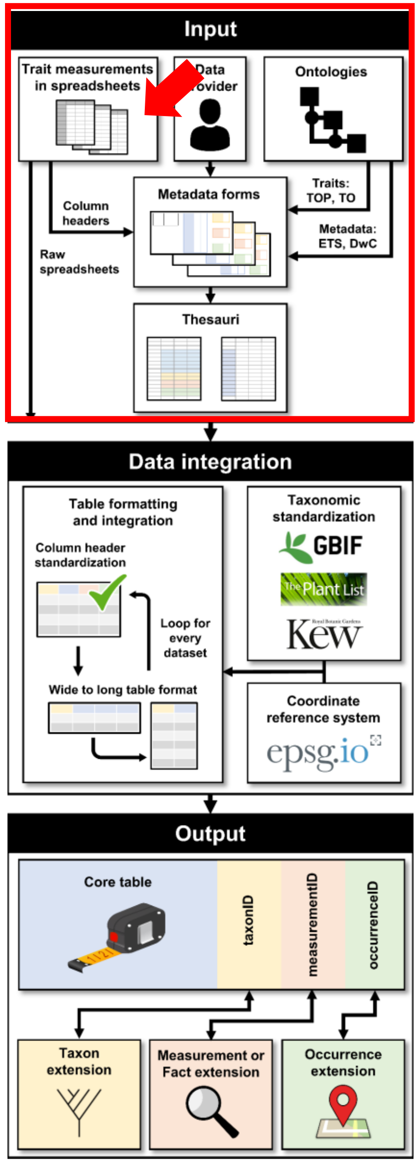
Herbari	Country	Plheigh	Stemhei	Stemdia	Sheathl
ny	panama		2		
ny	panama		4		
ny	panama	3		0.95	20
ny	panama		5	1.13	23
mo	panama	3		0.68	
mo	panama	2			
mo	panama	2		0.68	16.5
mo	panama	4		0.89	

Dlongitude	Elevation	Sex	Stemheight	Stemdiameter	Petiole
102.4	300		0.8	0.52	15
114.18		pistillate			
		pistillate		0.4	8.7
		sterile			5.7
		pistillate			29.5
114.13		staminate			
114.16		pistillate		0.87	10
105.21		pistillate			
108.31	372	sterile	1.5		45
106.25		pistillate			32.5
105.21		staminate			

Capturing trait measurements from the field or herbarium specimens



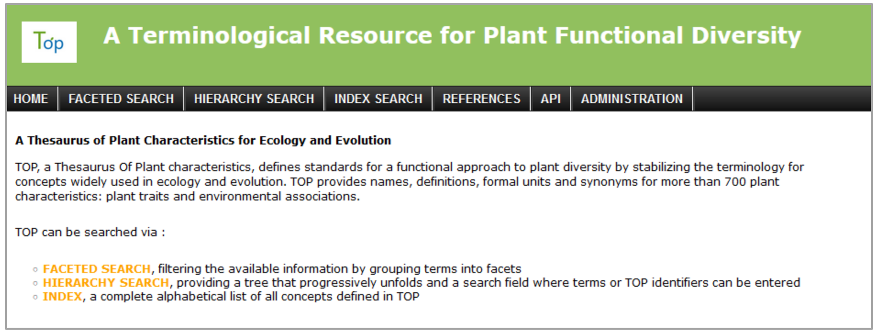


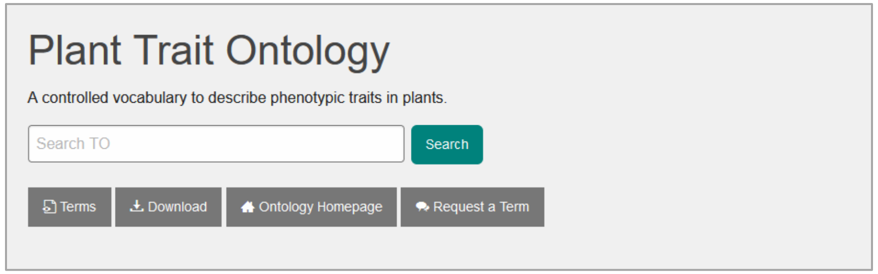


Lenters et al. 2021 *Ecological Informatics*

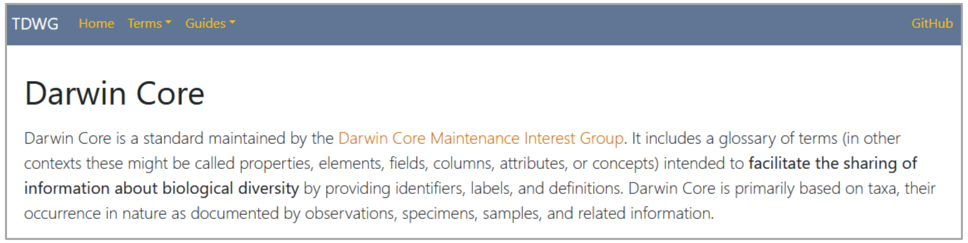
Workflow 1: Data integration & harmonization

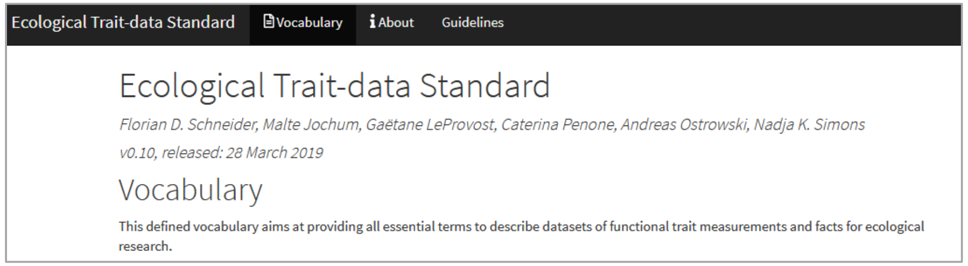
Ontologies for traits

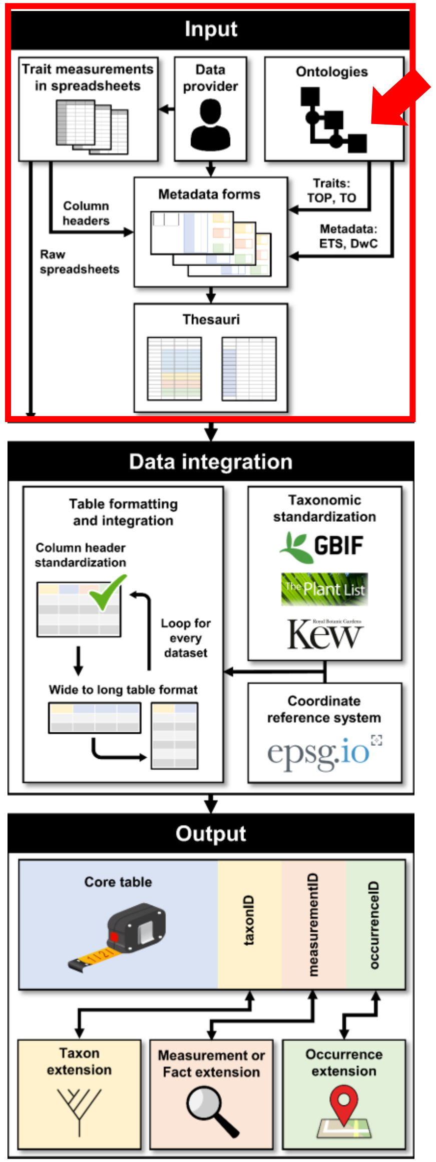




Additional metadata standards







➢ standardized trait names, metadata information and unambiguous identifiers (URI) for each term

Lenters et al. 2021 *Ecological Informatics*

Workflow 1: Data integration & harmonization

Metadata form

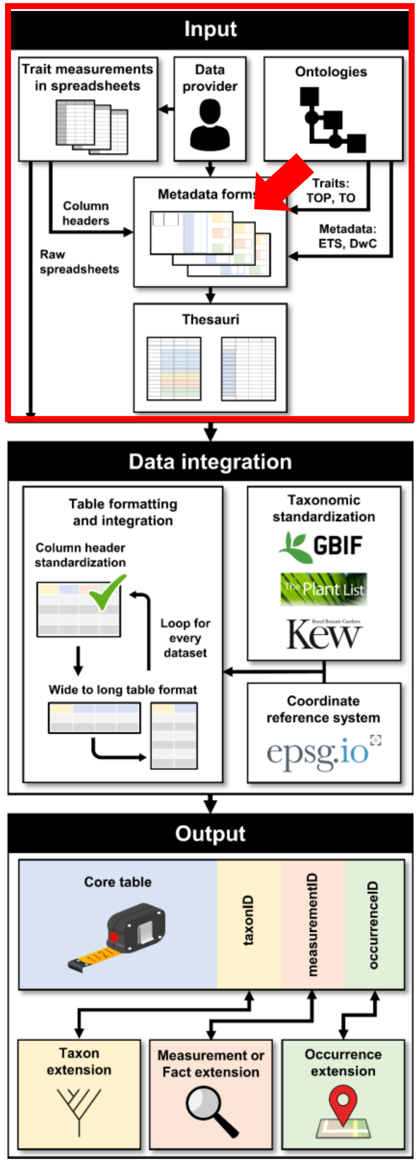
Name (dataset provider):	Guilherme Elias
Date (dd/mm/yyyy):	20/04/2020
Dataset name:	Arecaceae#1
Reference:	Elias, G. A., Colares, R., Antu...
verbatimCoordinateSystem:	☐ Decimal degrees ☐ Degrees-minutes-seconds ☑ UTM
verbatimSRS:	SIRGAS 2000 / UTM zone 22S
basisOfRecord: ? ? ? ? ?	◉ LivingSpecimen ○ PreservedSpecimen ○ FossilSpecimen ○ HumanObservation ○ MachineObservation

Quantitative traits	Fill in ⇩ (name)	Fill in ⇩ (unit)
Plant_height	H	m
Stem_length		
Stem_width	Tree Girth	cm
Leaf_sheat_length		
Petiole_length		
Petiole_thickness		
Peduncle_length		
Peduncle_width		
Distance_scar_bracteole		
Prophyll_length		
Peduncle_bract_length		
Apical_leaflet_length		
Apical_leaflet_width		
Apical_leaflet_angle		
Apical_leaflet_vein_count		
Basal_leaflet_angle		
Basal_leaflet_length		
Basal_leaflet_width		
Median_leaflet_length		
Median_leaflet_width		
Leaflet_number		

Taxon	Fill in ⇩
scientificName	Species
genus	
specificEpithet	
infraspecificEpithet	
originalNameUsage	
morphotype	
verbatimTaxonRank	

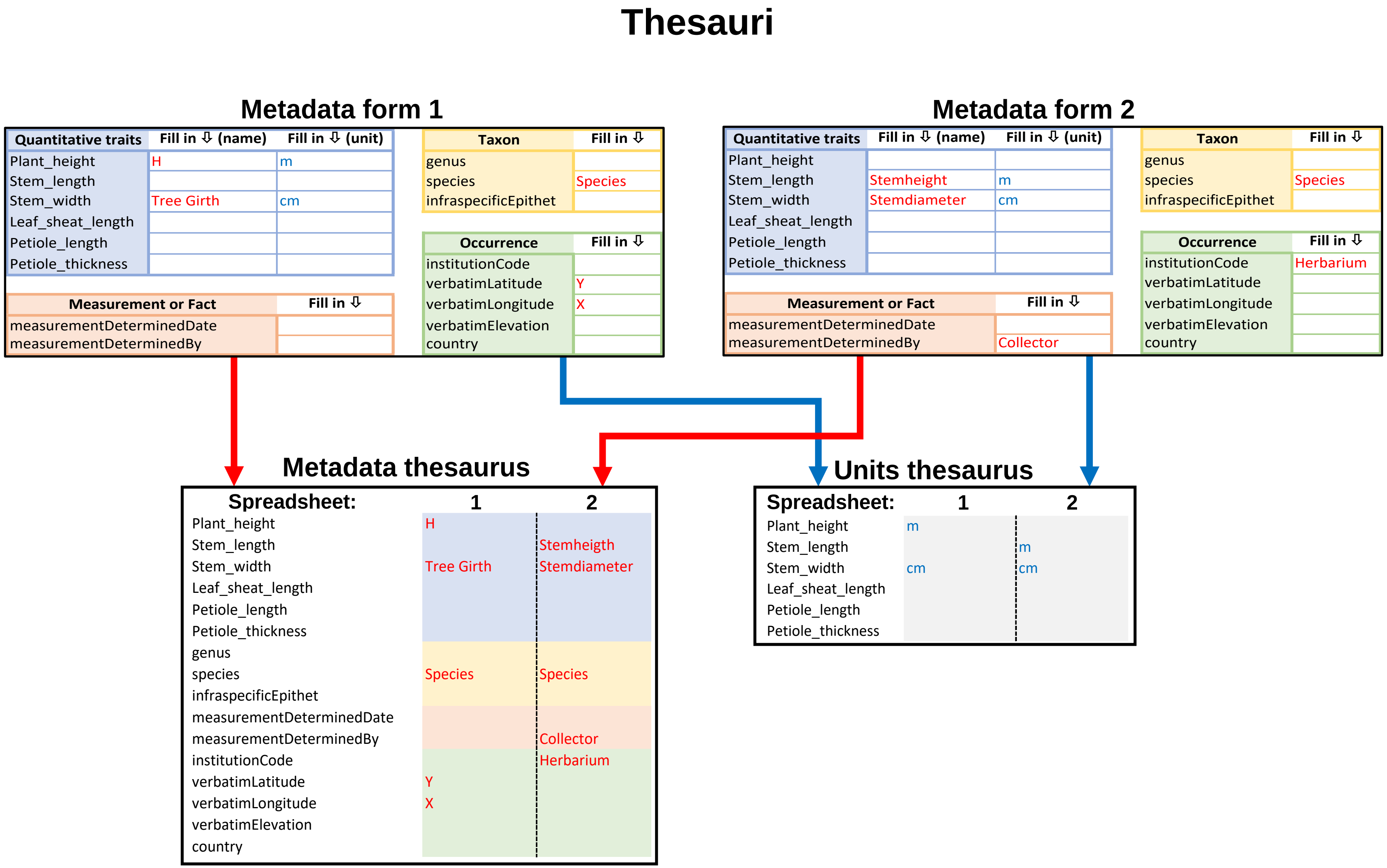
Measurement or Fact	Fill in ⇩
measurementDeterminedDate	
measurementDeterminedBy	

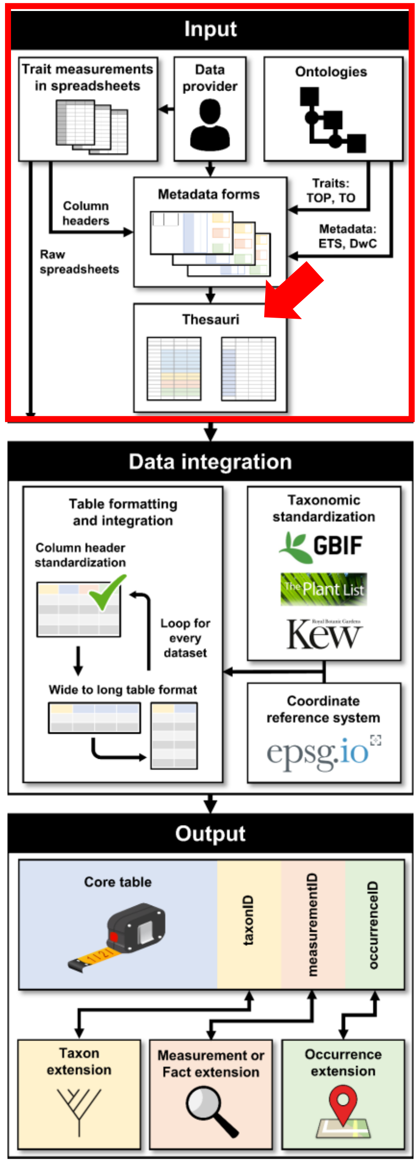
Occurrence	Fill in ⇩
identificationID	
recordNumber	
institutionCode	
verbatimLatitude	Y
verbatimLongitude	X
verbatimElevation	
country	
stateProvince	



Lenters et al. 2021 *Ecological Informatics*

Workflow 1: Data integration & harmonization





Lenters et al. 2021 *Ecological Informatics*

Workflow 1: Data integration & harmonization

Core table with standardized trait information

scientificName	verbatimTraitName	traitName	traitValue	traitUnit	traitID	taxonID	measurementID	occurrenceID
Syagrus romanzoffiana	Tree Girth	Stem_width	76	cm	http://purl...	1	NA	1
Syagrus romanzoffiana	Tree Girth	Stem_width	60	cm	http://purl...	1	NA	1
Syagrus romanzoffiana	Tree Girth	Stem_width	87	cm	http://purl...	1	NA	2
Syagrus romanzoffiana	H	Plant_height	1300	cm	http://purl...	1	NA	2
Syagrus romanzoffiana	H	Plant_height	800	cm	http://purl...	1	NA	2
Syagrus romanzoffiana	H	Plant_height	1300	cm	http://purl...	1	NA	2
Rhapis evansi	Stemheight	Stem_length	80	cm	http://purl...	2	1	3
Rhapis evansi	Stemdiameter	Stem_width	0.4	cm	http://purl...	2	1	3
Rhapis excelsa	Stemheight	Stem_length	150	cm	http://purl...	3	2	4
Rhapis excelsa	Stemheight	Stem_length	250	cm	http://purl...	3	2	4
Rhapis excelsa	Stemdiameter	Stem_width	0.87	cm	http://purl...	3	3	5
Rhapis excelsa	Stemdiameter	Stem_width	1.36	cm	http://purl...	3	3	5

<u>Trait data mobilization:</u>

- 138,993 individual trait measurements for 50 standardized traits covering 551 unique palm species (22%)

Extension tables

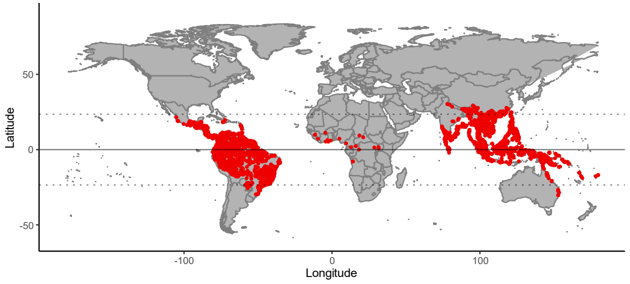
Taxon

GBIF

The Plant List

Royal Botanic Gardens Kew

Occurrence

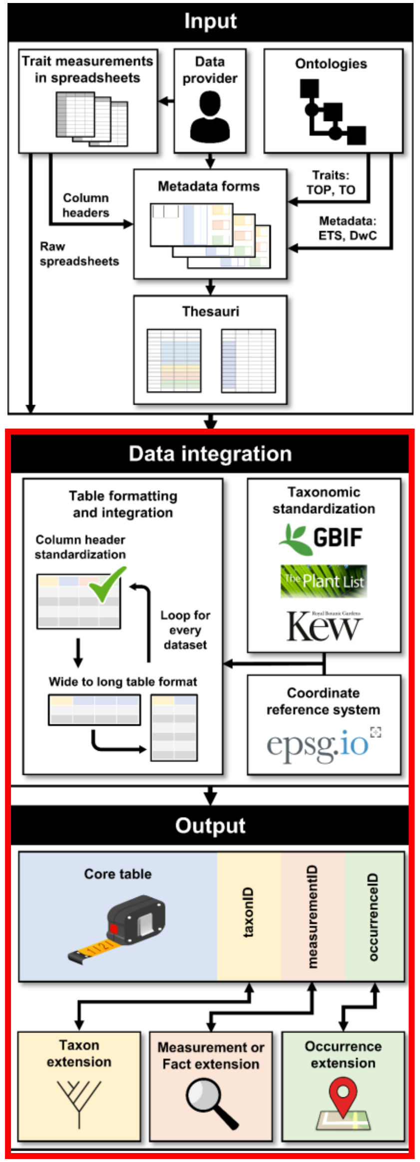


Measurement or Fact

measurementID

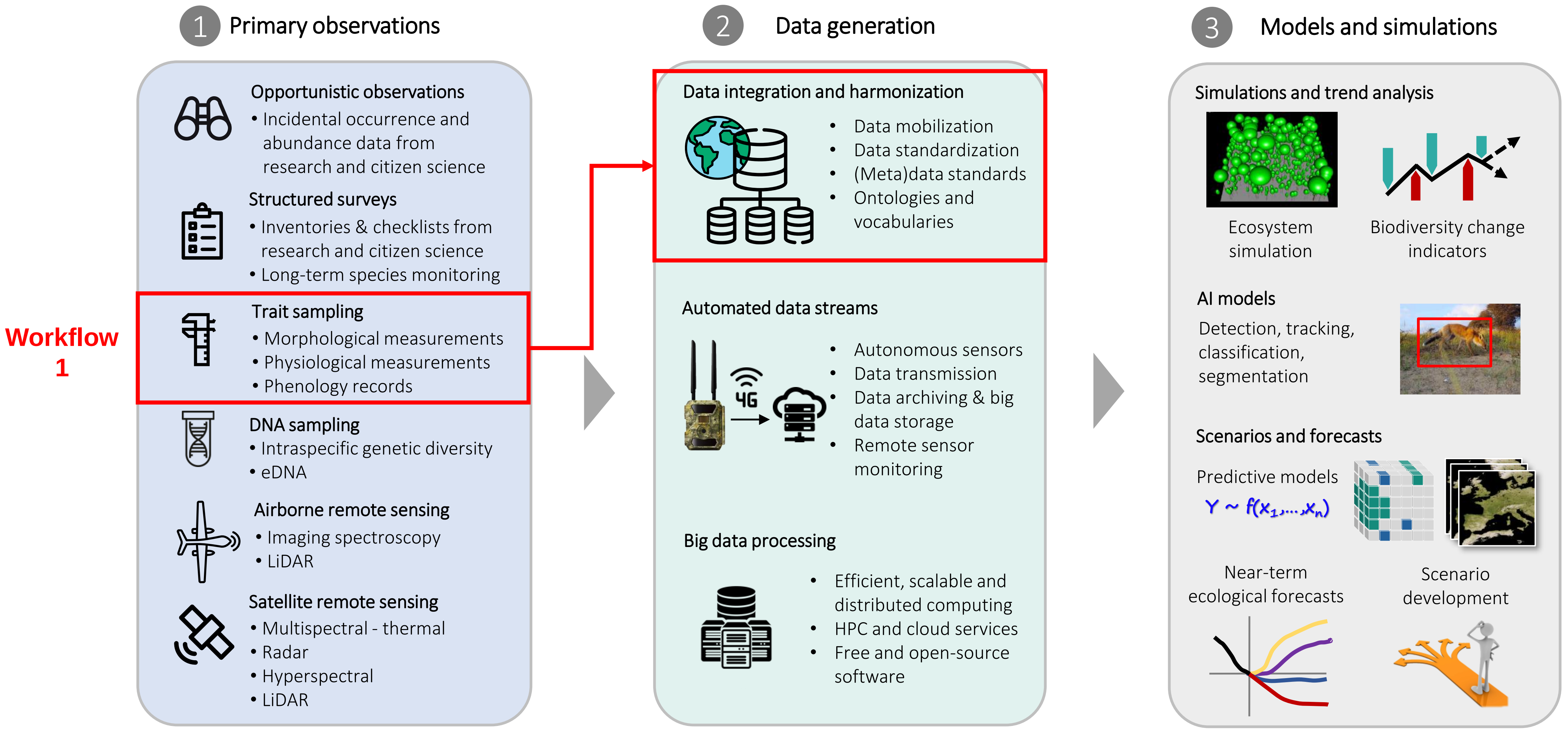
measurementDeterminedBy

basisOfRecord

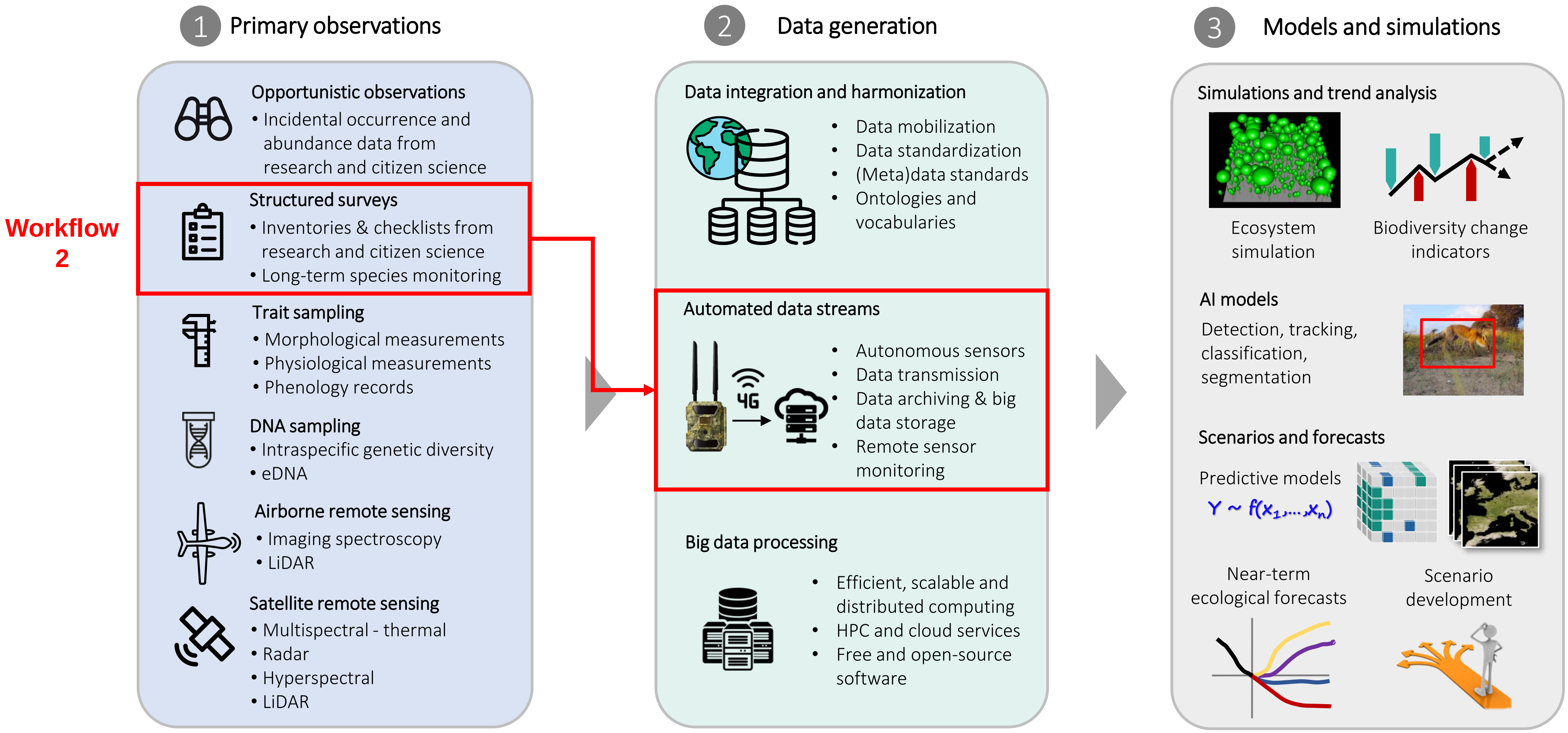


Lenters et al. 2021 *Ecological Informatics*

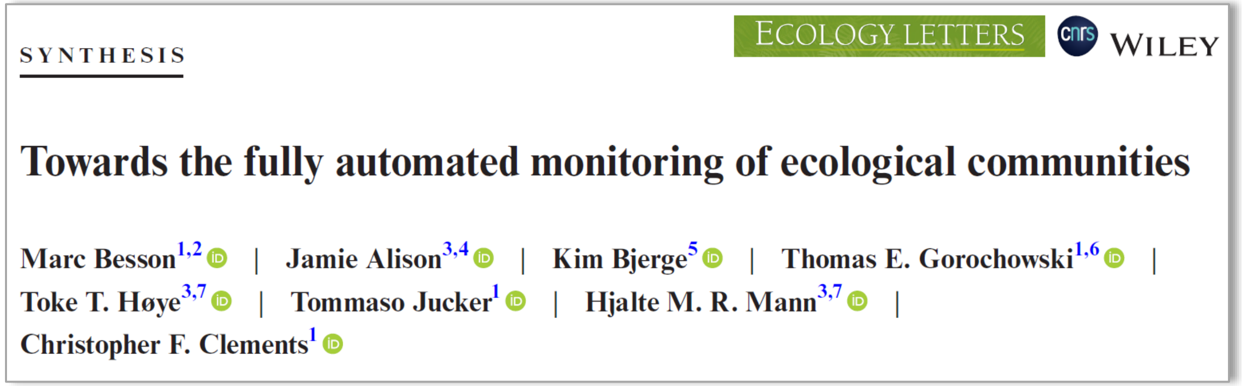
Biodiversity data



Biodiversity data



Workflow 2: Automated data streams



SYNTHESIS

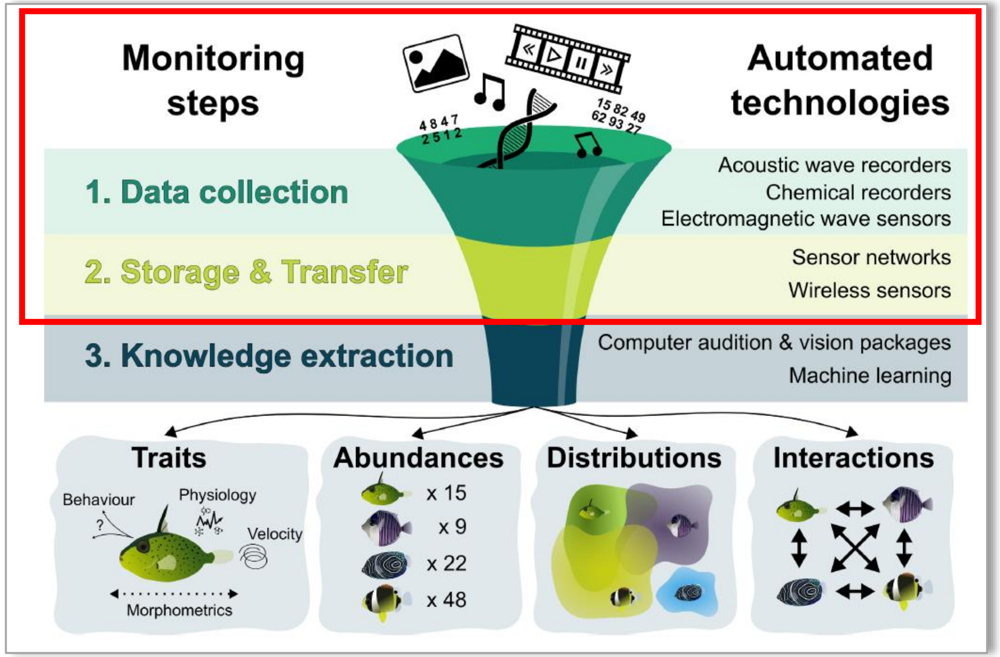
ECOLOGY LETTERS WILEY

Towards the fully automated monitoring of ecological communities

Marc Besson[1,2] | Jamie Alison[3,4] | Kim Bjerge[5] | Thomas E. Gorochowski[1,6] | Toke T. Høye[3,7] | Tommaso Jucker[1] | Hjalte M. R. Mann[3,7] | Christopher F. Clements[1]

Automated biodiversity monitoring:

- (near)real-time observations
- sampling of remote areas and at unfeasible times
- high-frequency observations without observer disturbance



Workflow 2: Automated data streams

Insect cameras	Automated Moth Trap	Wildlife cameras	AudioMoth	Bird radar	Aeroecology Cam
					

Autonomous deployment of sensors

- Power supply:
  - solar panels (12V/50-60W/100W) & batteries (AA or 12V)
  - main power (230 volts AC, 200-300 W)
- Data transmission:
  - 4G (SIM cards)
  - wired network

arise-biodiversity.nl

ARISE monitoring demonstration sites: https://www.arise-biodiversity.nl/teammonitoringdemonstration

Workflow 2: Automated data streams

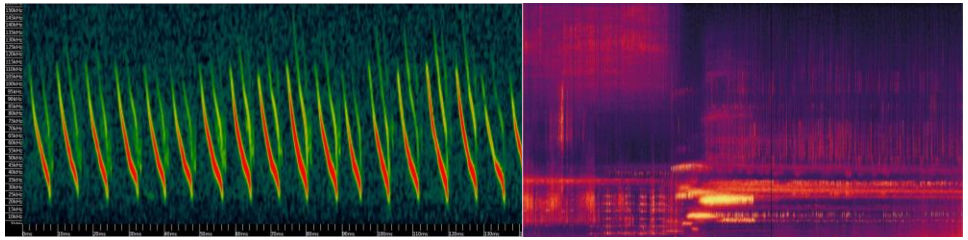
Number of devices | **Data volume**


- 10 × DIOPSIS insect cameras

- ~ 2.5 months/device
- ~ 1 Mio images
- ~ 715 GB

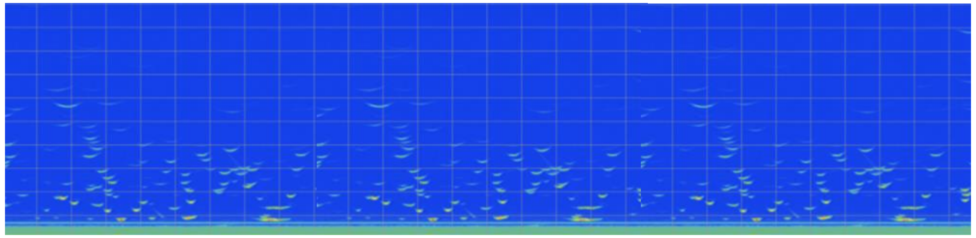

- 47 × wildlife cameras

- ~ 6 months/device
- ~ 770,000 images
- ~ 452 GB of data

- 26 × AudioMoth

- ~ 5.5 months/device
- ~ 6,000 audio files
- ~ 16 TB of data (compressed)

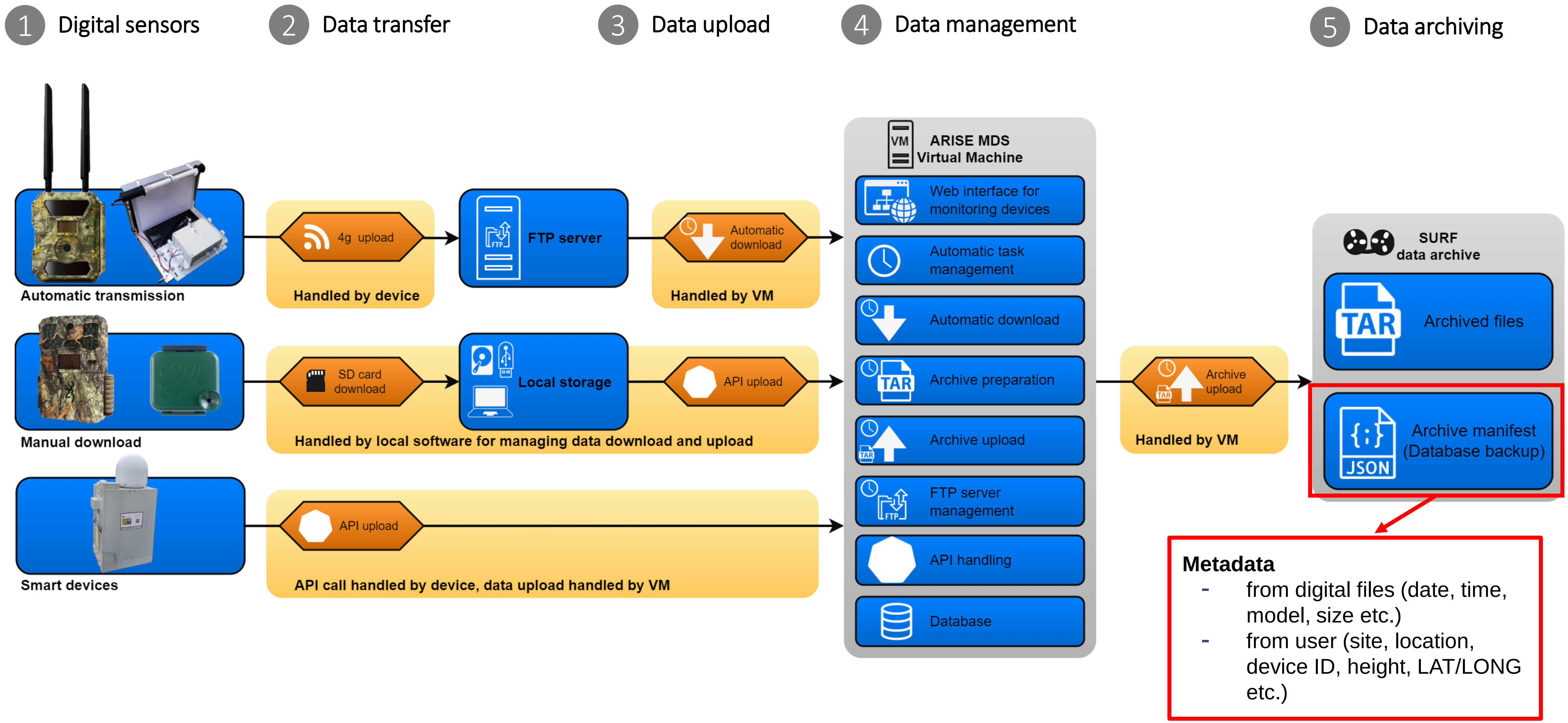



- 1 × Birdscan MR1 Radar

- ~ 1 year
- ~ 26 TB raw data
- ~ 0.5 TB derived data (signatures, thumbnails, PDP + MSSQL database)

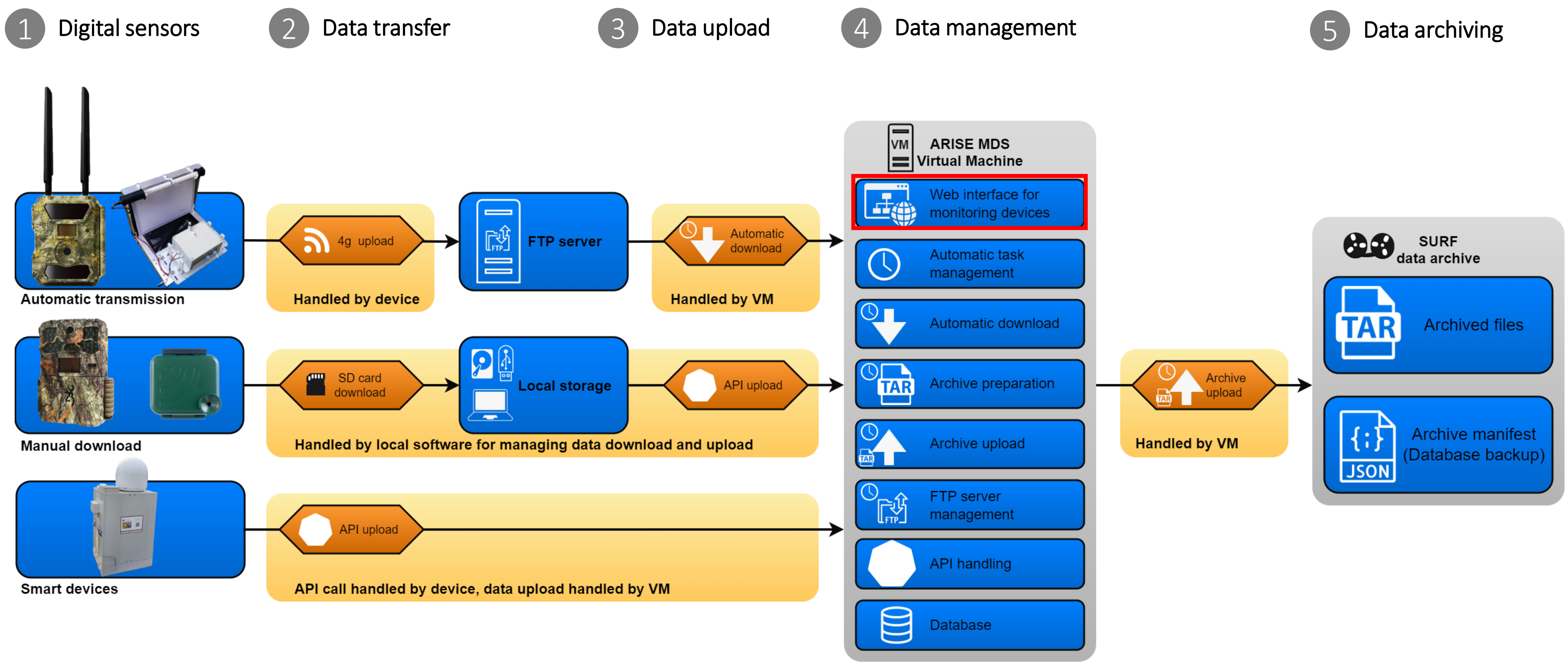
ARISE monitoring demonstration sites: https://www.arise-biodiversity.nl/teammonitoringdemonstration

Workflow 2: Automated data streams



ARISE monitoring demonstration sites: https://www.arise-biodiversity.nl/teammonitoringdemonstration

Workflow 2: Automated data streams

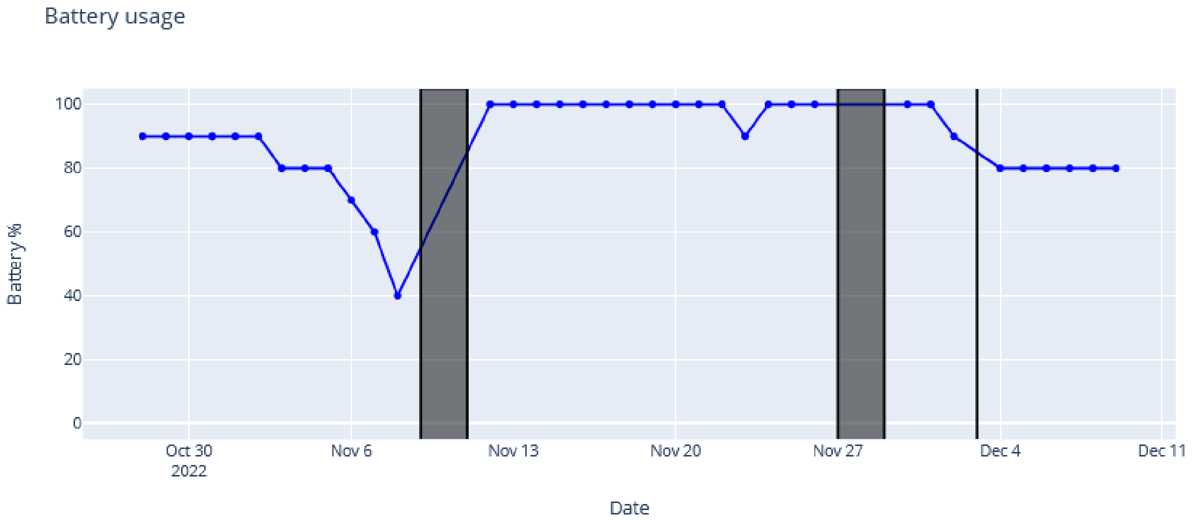


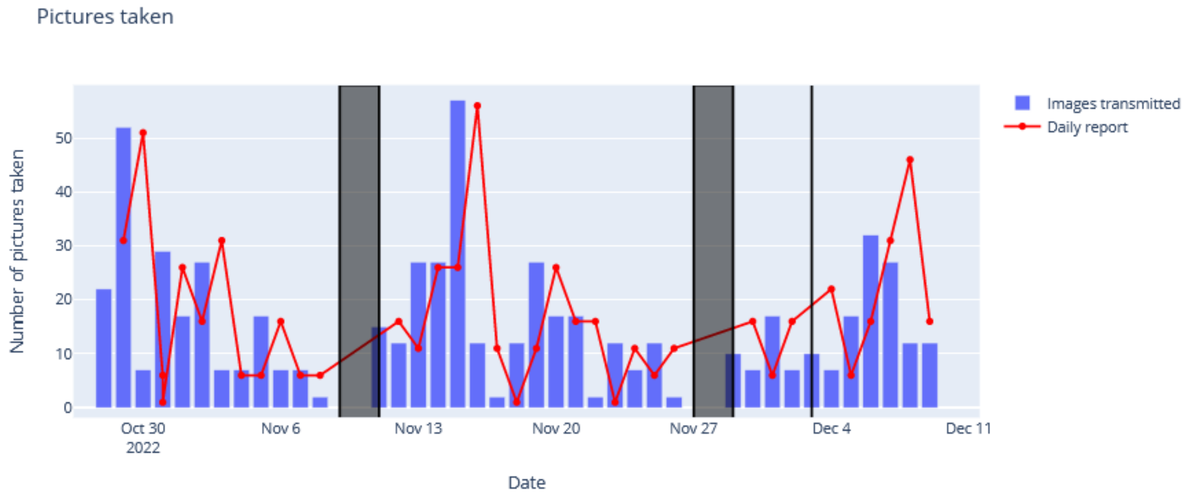
ARISE monitoring demonstration sites: https://www.arise-biodiversity.nl/teammonitoringdemonstration

Workflow 2: Automated data streams

(Near)real-time data viewing

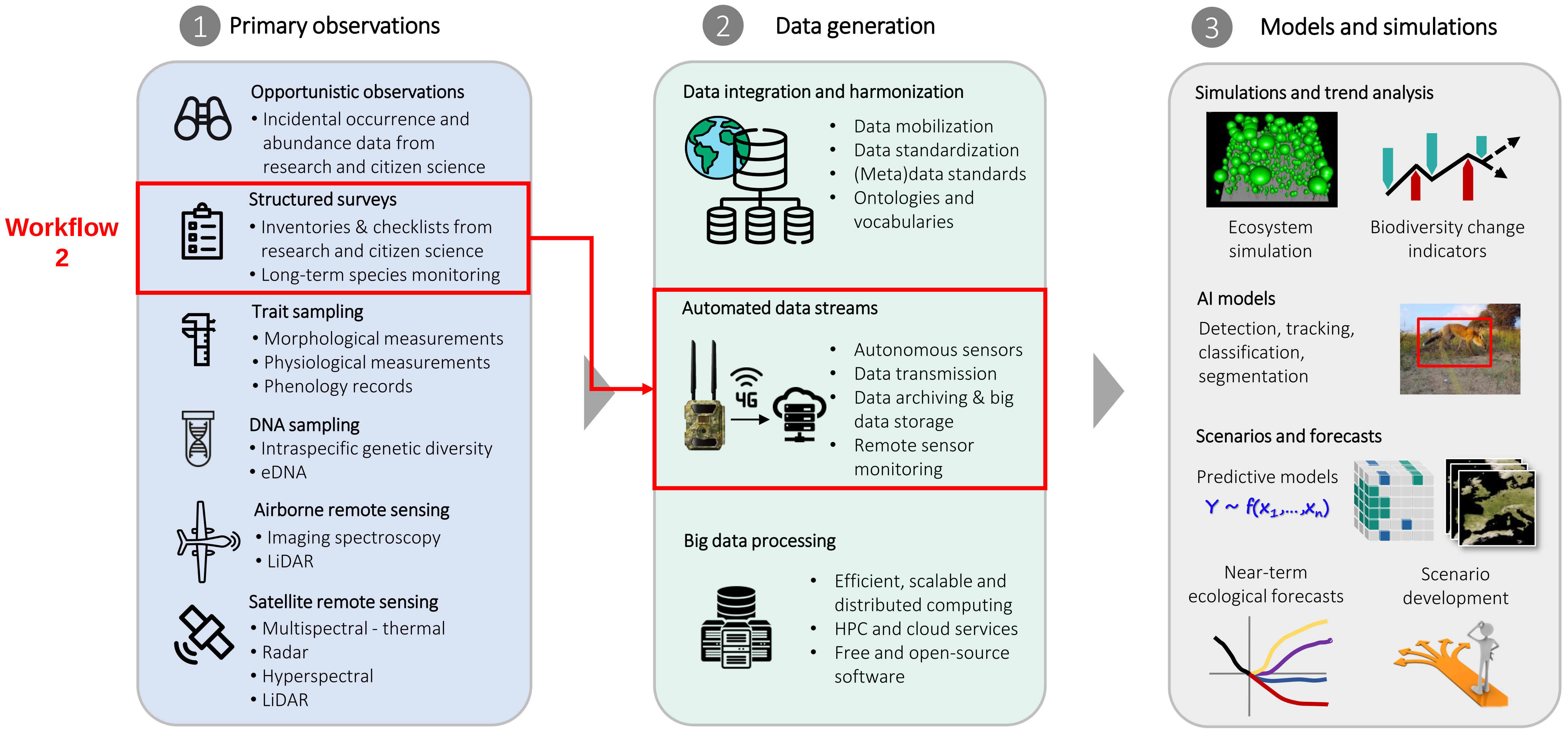
Remote monitoring of sensor performance





ARISE monitoring demonstration sites: https://www.arise-biodiversity.nl/teammonitoringdemonstration

Biodiversity data



Biodiversity data

1 Primary observations

Opportunistic observations
- Incidental occurrence and abundance data from research and citizen science

Structured surveys
- Inventories & checklists from research and citizen science
- Long-term species monitoring

Trait sampling
- Morphological measurements
- Physiological measurements
- Phenology records

DNA sampling
- Intraspecific genetic diversity
- eDNA

Workflow 3

Airborne remote sensing
- Imaging spectroscopy
- LiDAR

Satellite remote sensing
- Multispectral - thermal
- Radar
- Hyperspectral
- LiDAR

2 Data generation

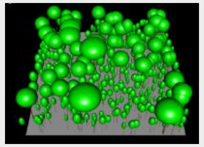
Data integration and harmonization
- Data mobilization
- Data standardization
- (Meta)data standards
- Ontologies and vocabularies

Automated data streams
- Autonomous sensors
- Data transmission
- Data archiving & big data storage
- Remote sensor monitoring

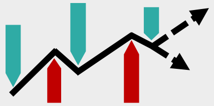
Big data processing
- Efficient, scalable and distributed computing
- HPC and cloud services
- Free and open-source software

3 Models and simulations

Simulations and trend analysis



Ecosystem simulation



Biodiversity change indicators

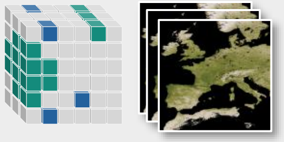
AI models

Detection, tracking, classification, segmentation

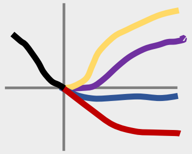
Scenarios and forecasts

Predictive models

$Y \sim f(x_1,\ldots,x_n)$

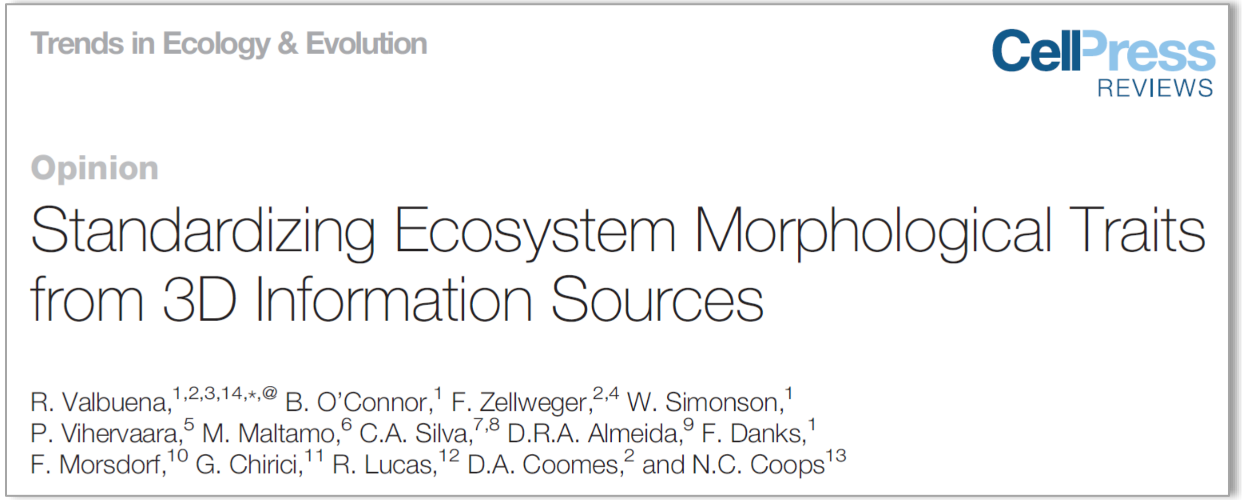


Near-term ecological forecasts



Scenario development

Workflow 3: Big data processing


Trends in Ecology & Evolution

CellPress REVIEWS

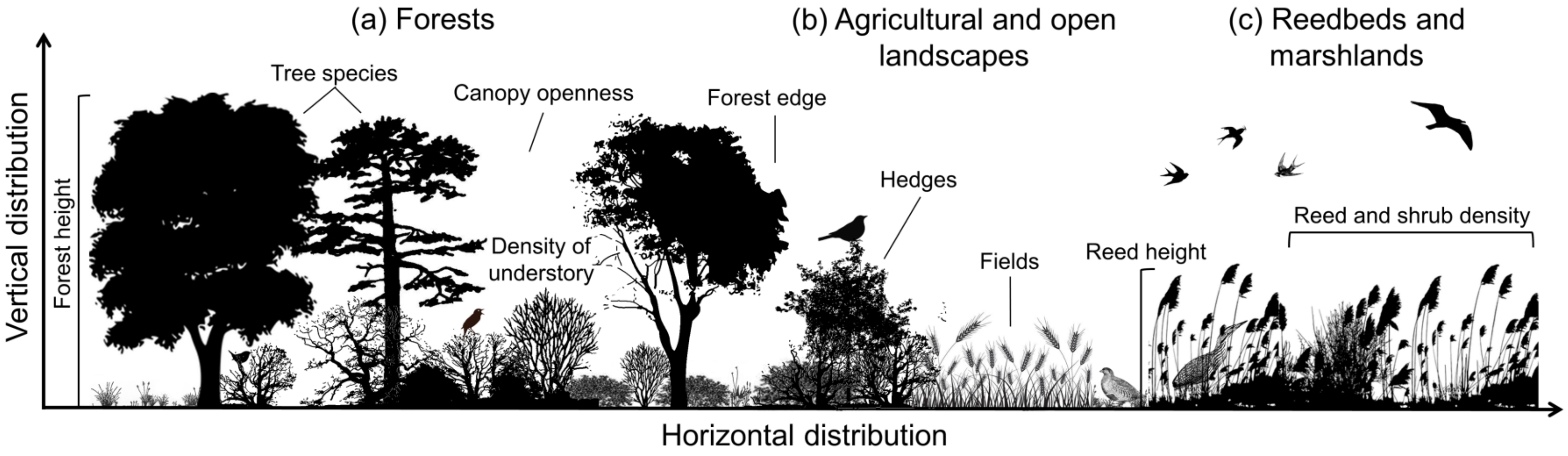
Opinion

Standardizing Ecosystem Morphological Traits from 3D Information Sources

R. Valbuena,[1,2,3,14,*,@] B. O'Connor,[1] F. Zellweger,[2,4] W. Simonson,[1] P. Vihervaara,[5] M. Maltamo,[6] C.A. Silva,[7,8] D.R.A. Almeida,[9] F. Danks,[1] F. Morsdorf,[10] G. Chirici,[11] R. Lucas,[12] D.A. Coomes,[2] and N.C. Coops[13]

Remote sensing data with 3D information:

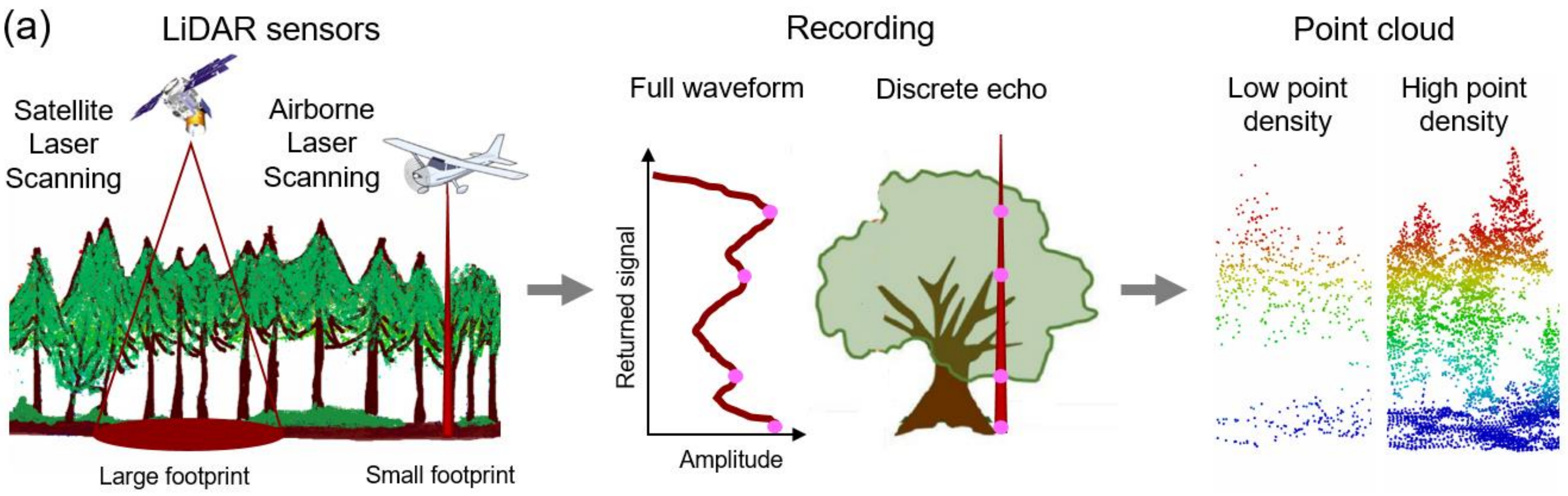
- large data volumes (e.g. multiple terabytes of raw data)
- many software tools are not open-access
- processing requires a high degree of specialization
- very important because they provide 3D measurements of ecosystem structure



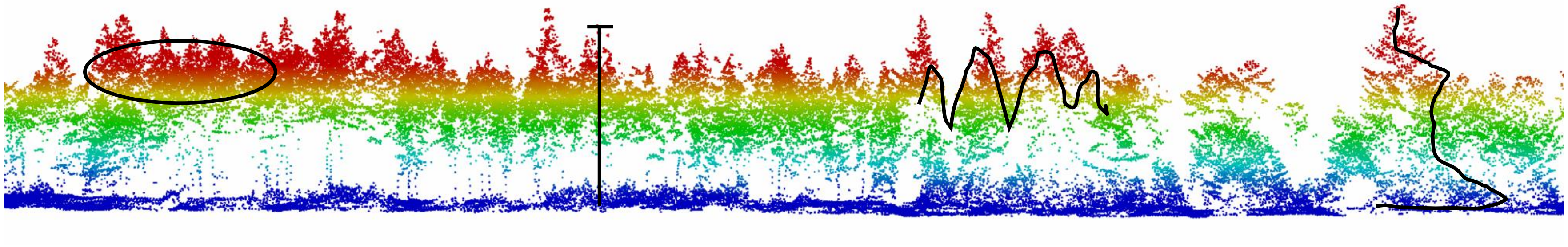
Kissling et al. 2017 *Research Ideas and Outcomes*

Workflow 3: Big data processing

LiDAR
Light Detection And Ranging (LiDAR) technology provides 3D point clouds of vegetation, terrain and infrastructure from which metrics of ecosystem structure can be derived.



LiDAR metrics
Statistical properties of 3D point clouds characterizing vegetation structure



Bakx et al. 2019 *Diversity and Distributions*

Workflow 3: Big data processing

SoftwareX 12 (2020) 100626

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

ELSEVIER

Original software publication

Laserchicken—A tool for distributed feature calculation from massive LiDAR point cloud datasets

C. Meijer [a,*], M.W. Grootes [a], Z. Koma [b], Y. Dzigan [a], R. Gonçalves [a], B. Andela [a], G. van den Oord [a], E. Ranguelova [a], N. Renaud [a], W.D. Kissling [b]

[a] Netherlands eScience Center, Science Park 140, 1098 XG Amsterdam, The Netherlands
[b] Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, P.O. Box 94240, 1090 GE Amsterdam, The Netherlands

Laserchicken software:

- cross-platform Python tool for extracting statistical properties of 3D point clouds
- free and open-source software (FOSS)
- designed for efficient, scalable, distributed processing of multi-terabyte datasets

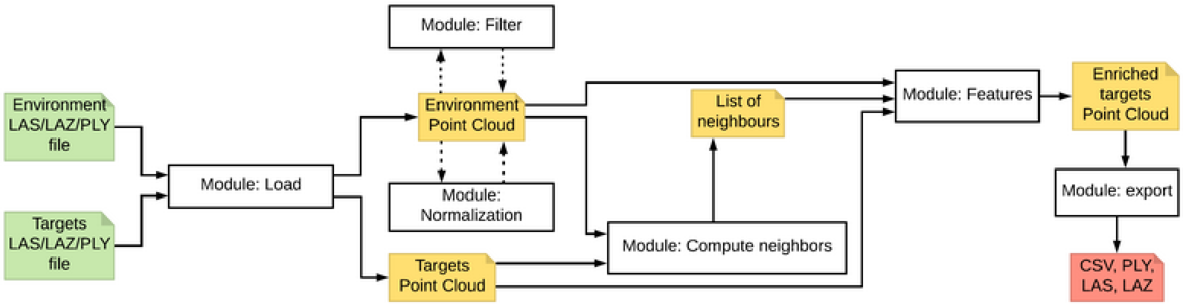
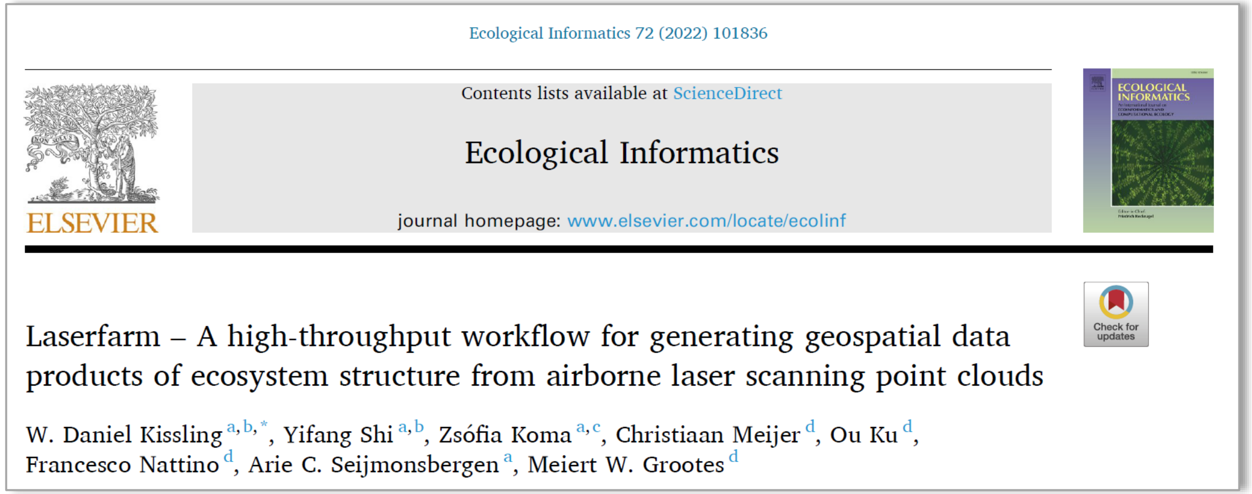


Table 1
Features currently implemented in Laserchicken..

Feature name	Formal description	Example of use	References
Point density	$\frac{N}{V}$ where V is the target volume or area	Point cloud spatial distribution	
Pulse penetration ratio	$\frac{N_{\text{ground}}}{N_{\text{total}}}$	Tree species classification	[17]
Echo ratio	$100 \cdot \frac{N_{3D}}{N_{2D}}$	Roof detection	[18]
Skewness	$\frac{1}{\sigma^3} \cdot \sum \frac{(Z_i-\bar{Z})^3}{N}$	Vegetation, ground, and roof classification and detection	[19]
Kurtosis	$\frac{1}{\sigma^4} \cdot \sum \frac{(Z_i-\bar{Z})^4}{N}$	Vegetation, ground, and roof classification and detection	[19]
Standard deviation	$\sqrt{\sum \frac{(Z_i-\bar{Z})^2}{N-1}}$	Classification of reed within wetland	[20]
Variance	$\sum \frac{(Z_i-\bar{Z})^2}{N-1}$	Classification of reed within wetland	[20]
Sigma Z	$\sqrt{\sum \frac{(R_i-\bar{R})^2}{N-1}}$ where R_i is the residual after plane fitting		Adapted from [20]
Minimum Z	Z_{min}	Simple digital terrain model in wetlands	[20]
Maximum Z	Z_{max}	Height and structure of forests	[21]
Mean Z	$\frac{1}{N} \cdot \sum Z_i$	Height and structure of forests	[21]
Median Z	Z_{median}	Height and structure of forests	[21]
Range Z	$\lvert Z_{\text{max}} - Z_{\text{min}} \rvert$	Height and structure of forests	[21]
Percentiles Z	Height of every 10^{th} percentile.	Height and structure of forests	[21]
Eigenvalues	$\lambda_1, \lambda_2, \lambda_3$, with $\lvert\lambda_1\rvert \geq \lvert\lambda_2\rvert \geq \lvert\lambda_3\rvert$	Classification of urban objects	[22]
Normal vector	eigen vector $\vec{v}_3$	Roof detection	[23]
Slope	$\tan(\arccos(\vec{v}_3 \cdot \vec{k}))$, where $\vec{k} = [0, 0, 1]^T$	Planar surface detection	[24]
Entropy Z	$-\sum_i P_i \cdot \log_2 P_i$, with $P_i = \frac{N_i}{\sum_j N_j}$ and N_i points in bin i	Foliage height diversity	[25]
Coefficient variance Z	$\frac{1}{\bar{Z}} \cdot \sqrt{\sum \frac{(Z_i-\bar{Z})^2}{N-1}}$	Urban tree species classification	[8]
Non-ground density absolute mean	$\frac{100}{N_{\text{non-ground}}} \cdot \sum_{i \in \text{non-ground}} [Z_i > \bar{Z}_{\text{non-ground}}]$	Urban tree species classification	[8]
Band ratio	$\frac{N_{Z_i < Z < Z_j}}{N_{\text{tot}}}$ with Z_i and Z_j provided by user	Height and vertical structure of vegetation	

Meijer et al. 2020 *SoftwareX*

Workflow 3: Big data processing


Ecological Informatics 72 (2022) 101836

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

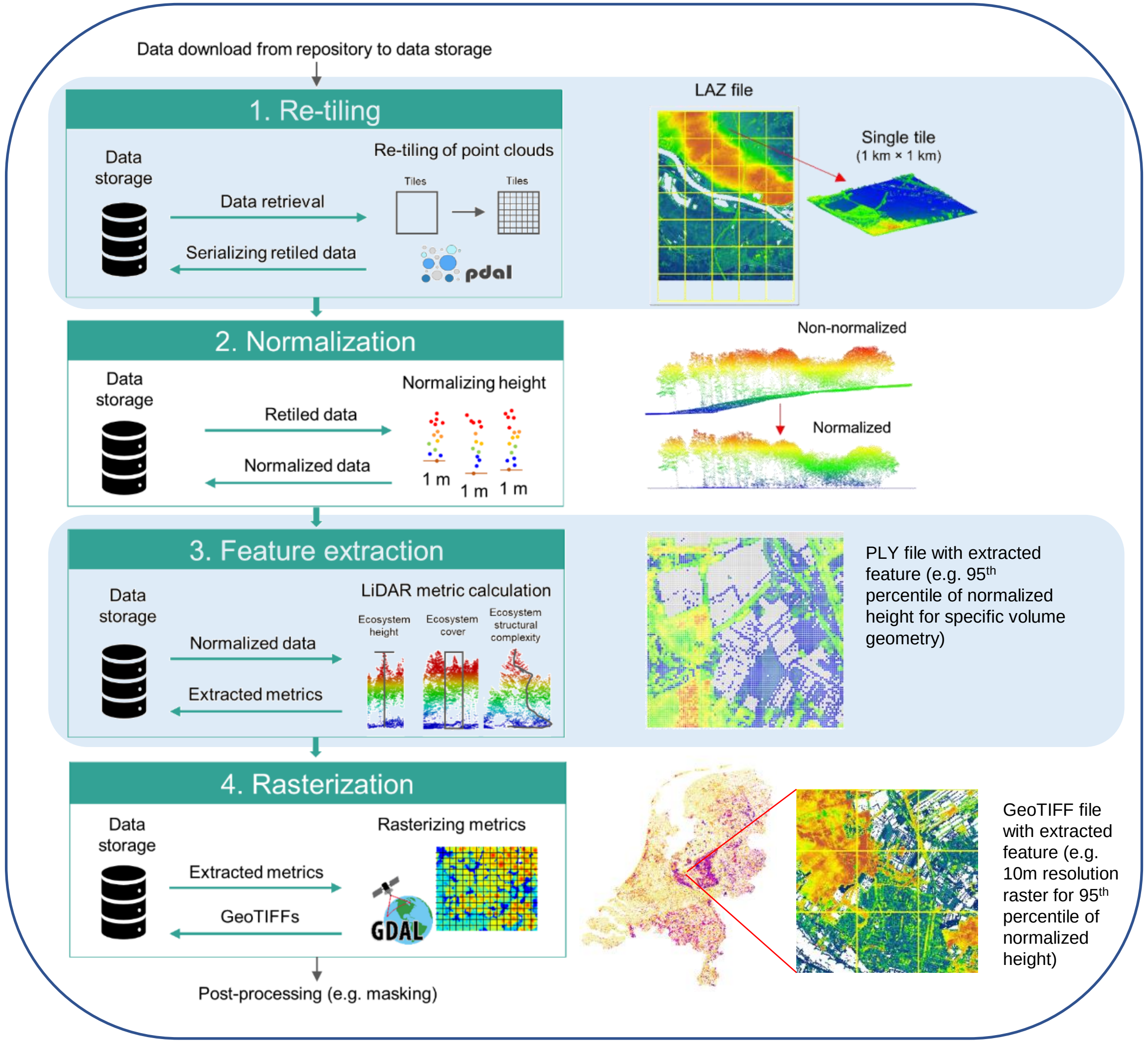
ELSEVIER

Laserfarm – A high-throughput workflow for generating geospatial data products of ecosystem structure from airborne laser scanning point clouds

W. Daniel Kissling [a,b,*], Yifang Shi [a,b], Zsófia Koma [a,c], Christiaan Meijer [d], Ou Ku [d], Francesco Nattino [d], Arie C. Seijmonsbergen [a], Meiert W. Grootes [d]

Laserfarm workflow:

- High Performance Computing (HPC) workflow for processing multi-terabyte LiDAR point clouds
- Deployable on various computing infrastructures (e.g. cluster of virtual machines, cloud computing environments)
- Free and open-source software (FOSS) tool
- Implemented in Python and available as Jupyter notebook



Kissling et al. 2022 *Ecological Informatics*

Workflow 3: Big data processing

Laserfarm workflow tested on multiple IT infrastructures and platforms

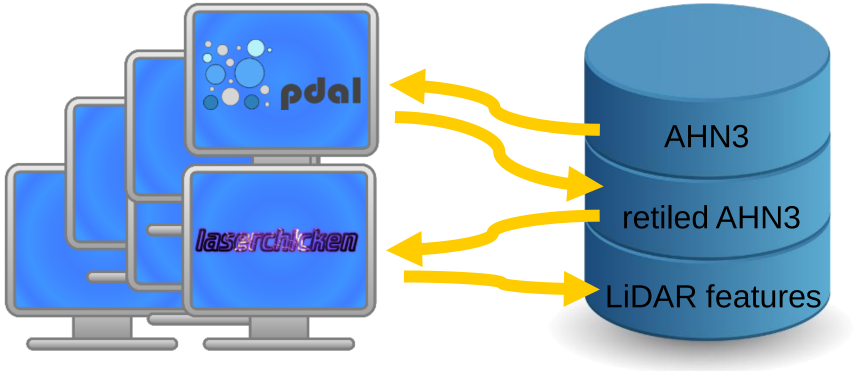
SURFsara HPC Cloud and Grid

https://www.surf.nl/en/about-surf/subsidiaries/surfsara/

HPC Cloud

- Cluster
- Multi-core processors for multi-threaded applications
- Fast network for distributed memory applications

Data storage (dCache)



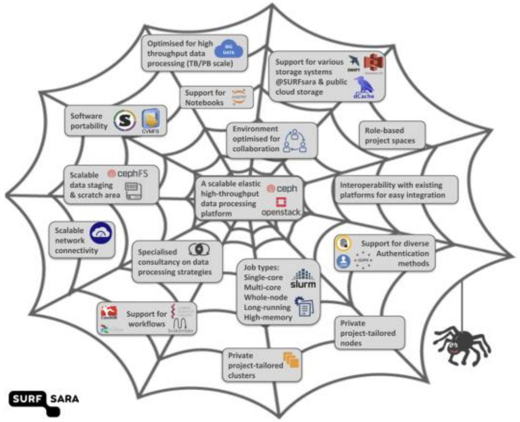
- Distributed data storage system
- Cluster of machines that mimics a large disc
- Data storage: 100 TB

Microsoft Azure Cloud

Microsoft Azure

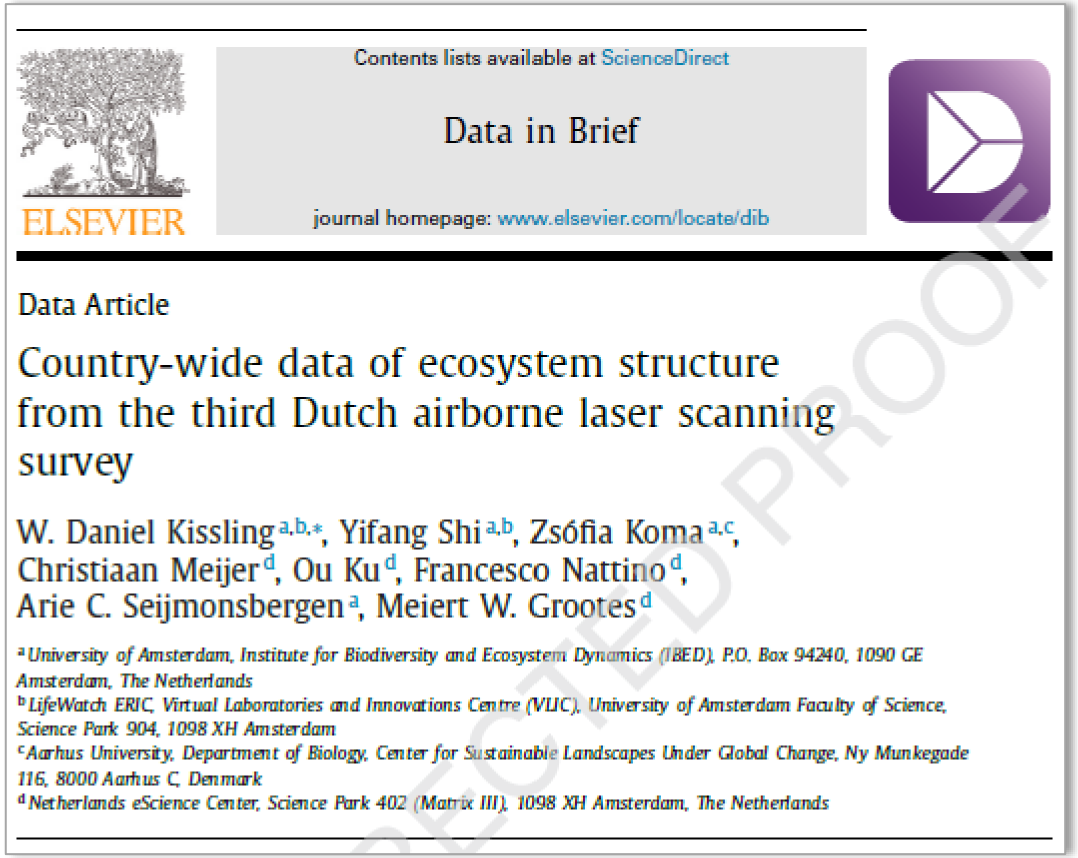
LifeWatch

SURFsara Spider



Kissling et al. 2022 *Ecological Informatics*

Workflow 3: Big data processing


Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

ELSEVIER

Data Article

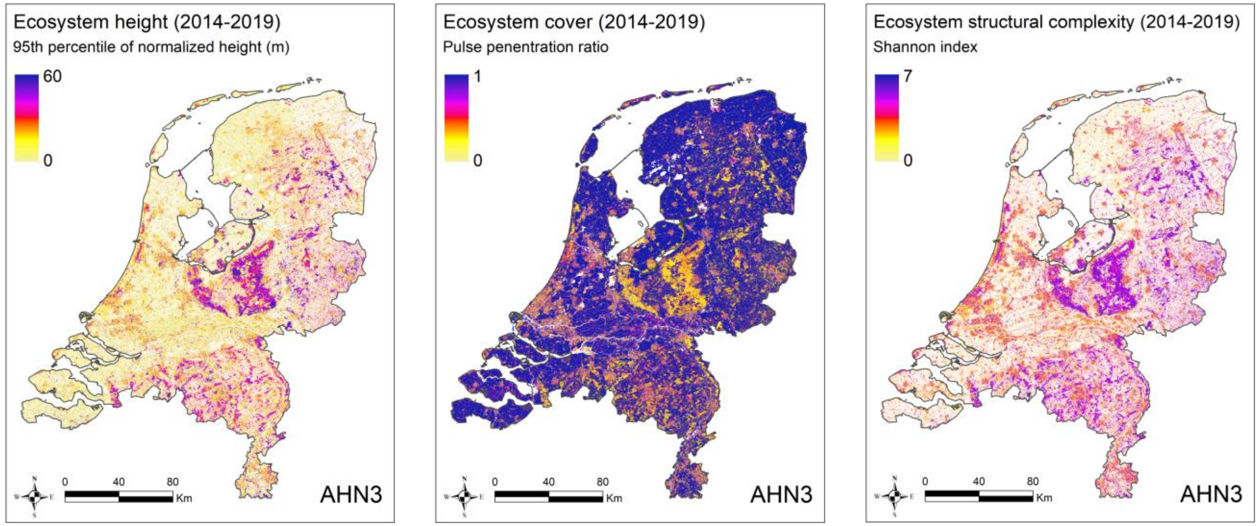
Country-wide data of ecosystem structure from the third Dutch airborne laser scanning survey

W. Daniel Kissling[a,b,*], Yifang Shi[a,b], Zsófia Koma[a,c], Christiaan Meijer[d], Ou Ku[d], Francesco Nattino[d], Arie C. Seijmonsbergen[a], Meiert W. Grootes[d]

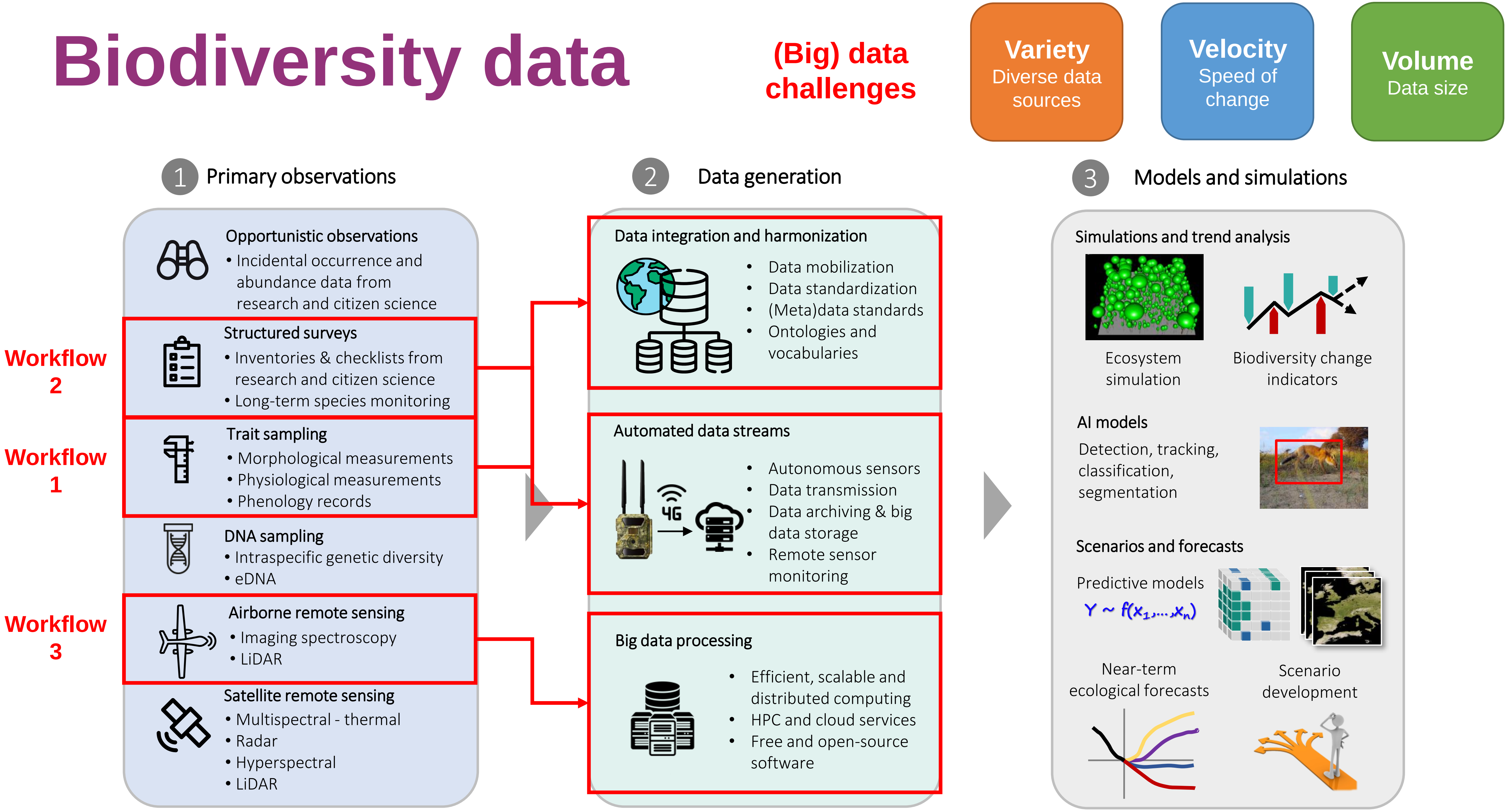
[a] University of Amsterdam, Institute for Biodiversity and Ecosystem Dynamics (IBED), P.O. Box 94240, 1090 GE Amsterdam, The Netherlands
[b] LifeWatch ERIC, Virtual Laboratories and Innovations Centre (VLIC), University of Amsterdam Faculty of Science, Science Park 904, 1098 XH Amsterdam
[c] Aarhus University, Department of Biology, Center for Sustainable Landscapes Under Global Change, Ny Munkegade 116, 8000 Aarhus C, Denmark
[d] Netherlands eScience Center, Science Park 402 (Matrix III), 1098 XH Amsterdam, The Netherlands

Data products of ecosystem structure:

- 10 m resolution raster layers of ecosystem structure
- 25 LiDAR metrics of ecosystem height, cover and structural complexity
- GeoTIFF files
- can be readily used by ecologists in a geographic information system (GIS) or analytical open-source software such as R



Kissling et al. 2022 *Data in Brief*



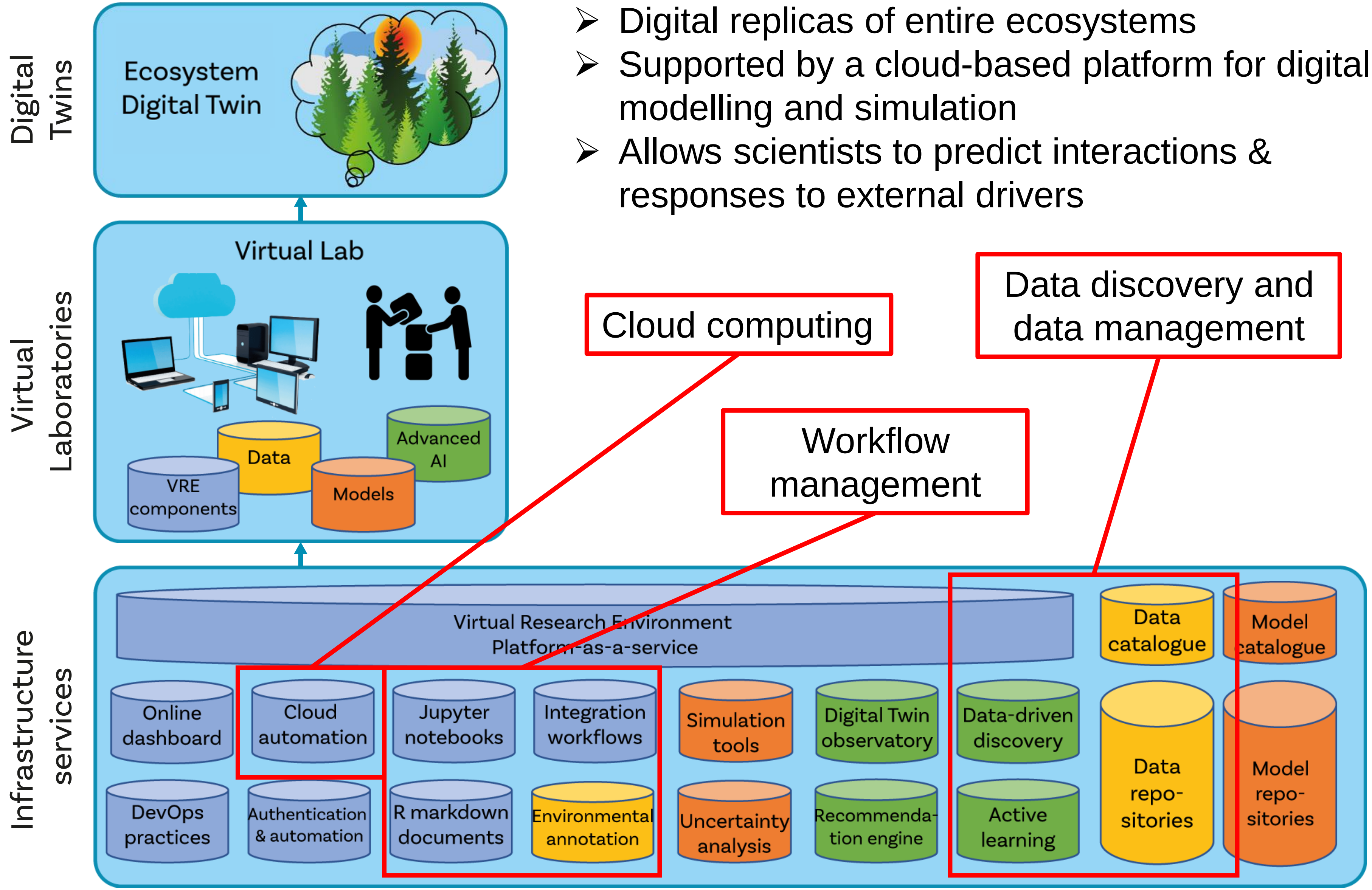
Biodiversity data
(Big) data challenges
Variety
Diverse data sources
Velocity
Speed of change
Volume
Data size
1 Primary observations
2 Data generation
3 Models and simulations
Opportunistic observations
• Incidental occurrence and abundance data from research and citizen science
Structured surveys
• Inventories & checklists from research and citizen science
• Long-term species monitoring
Trait sampling
• Morphological measurements
• Physiological measurements
• Phenology records
DNA sampling
• Intraspecific genetic diversity
• eDNA
Airborne remote sensing
• Imaging spectroscopy
• LiDAR
Satellite remote sensing
• Multispectral - thermal
• Radar
• Hyperspectral
• LiDAR
Workflow 2
Workflow 1
Workflow 3
Data integration and harmonization
• Data mobilization
• Data standardization
• (Meta)data standards
• Ontologies and vocabularies
Automated data streams
4G
• Autonomous sensors
• Data transmission
• Data archiving & big data storage
• Remote sensor monitoring
Big data processing
• Efficient, scalable and distributed computing
• HPC and cloud services
• Free and open-source software
Simulations and trend analysis
Ecosystem simulation
Biodiversity change indicators
AI models
Detection, tracking, classification, segmentation
Scenarios and forecasts
Predictive models
$Y \sim f(x_1, \ldots, x_n)$
Near-term ecological forecasts
Scenario development

Digital Twins of ecosystems

Simulation platform for complex (eco)system analysis

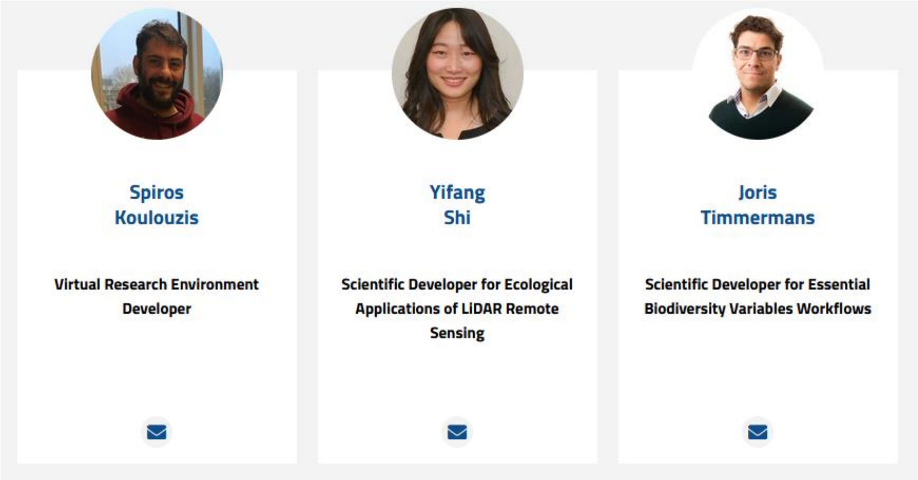
Interactive, collaborative, virtual research environment

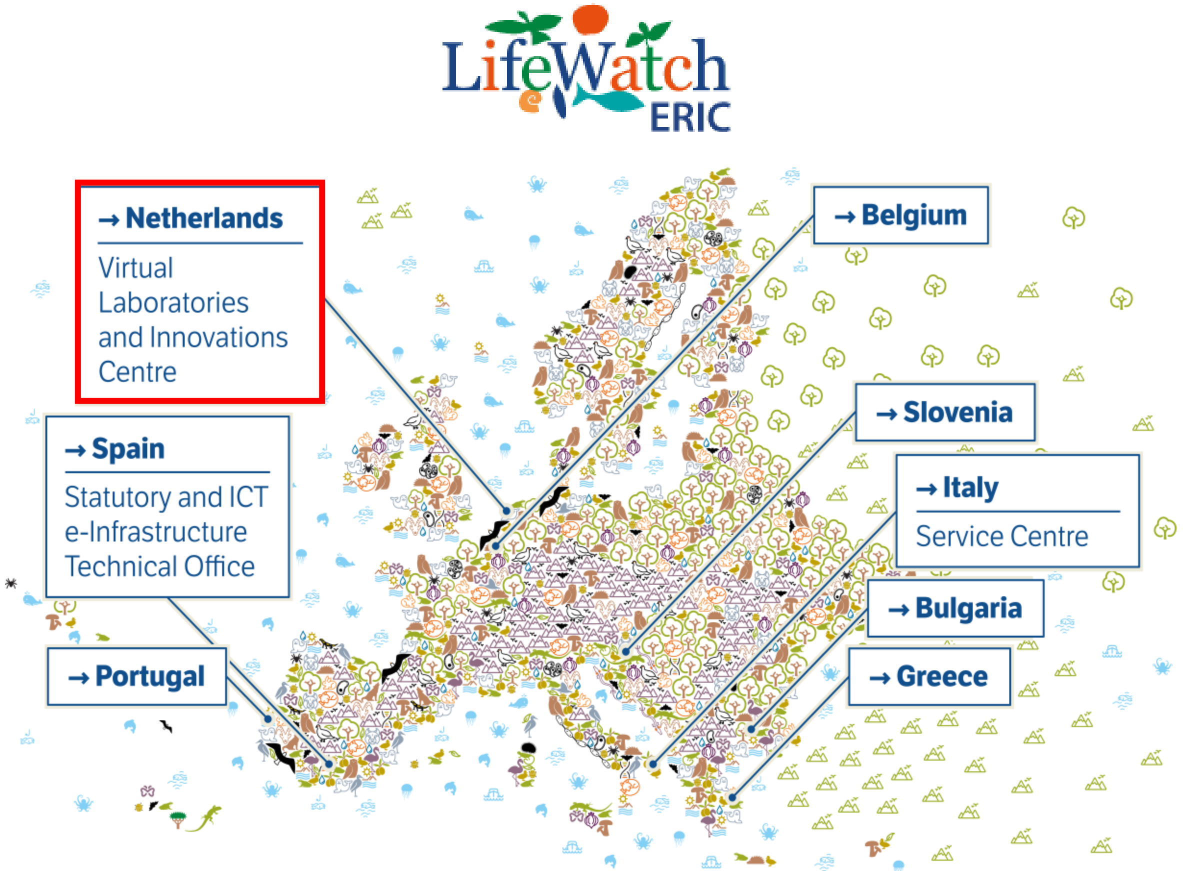
What do we need to develop Digital Twins of ecosystems?



- Digital replicas of entire ecosystems
- Supported by a cloud-based platform for digital modelling and simulation
- Allows scientists to predict interactions & responses to external drivers

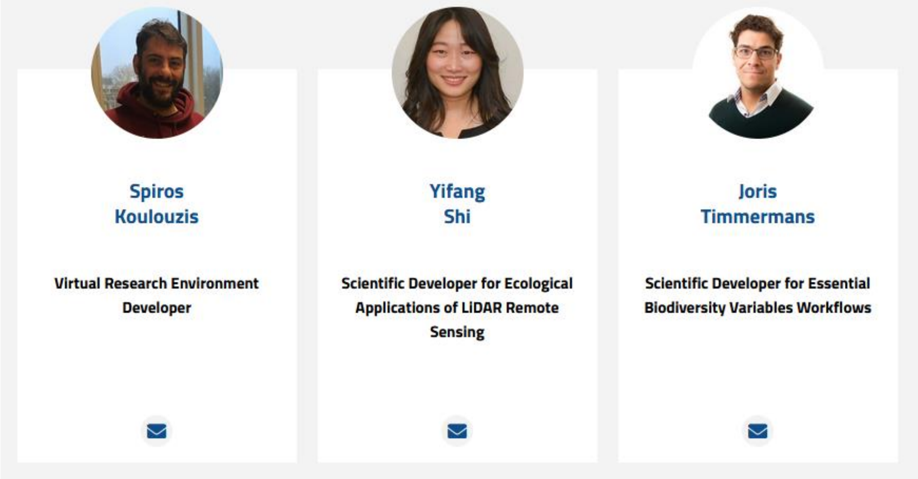
Virtual research environments





https://www.lifewatch.eu/organisation-governance/vlab-innovation-centre/

Virtual research environments



Received: 24 November 2021 | Revised: 21 April 2022 | Accepted: 3 May 2022
DOI: 10.1002/spe.3098

RESEARCH ARTICLE

WILEY

Notebook-as-a-VRE (NaaVRE): From private notebooks to a collaborative cloud virtual research environment

Zhiming Zhao[1,2] | Spiros Koulouzis[1,2] | Riccardo Bianchi[1,2] | Siamak Farshidi[1] | Zeshun Shi[1] | Ruyue Xin[1] | Yuandou Wang[1] | Na Li[1] | Yifang Shi[2,3] | Joris Timmermans[2,3] | W. Daniel Kissling[2,3]

Notebook-as-a-VRE:

- building a VRE by extending the notebook environment of end-users
- current design based on Jupyter, but other environments (e.g. R-Studio) can also be considered and supported

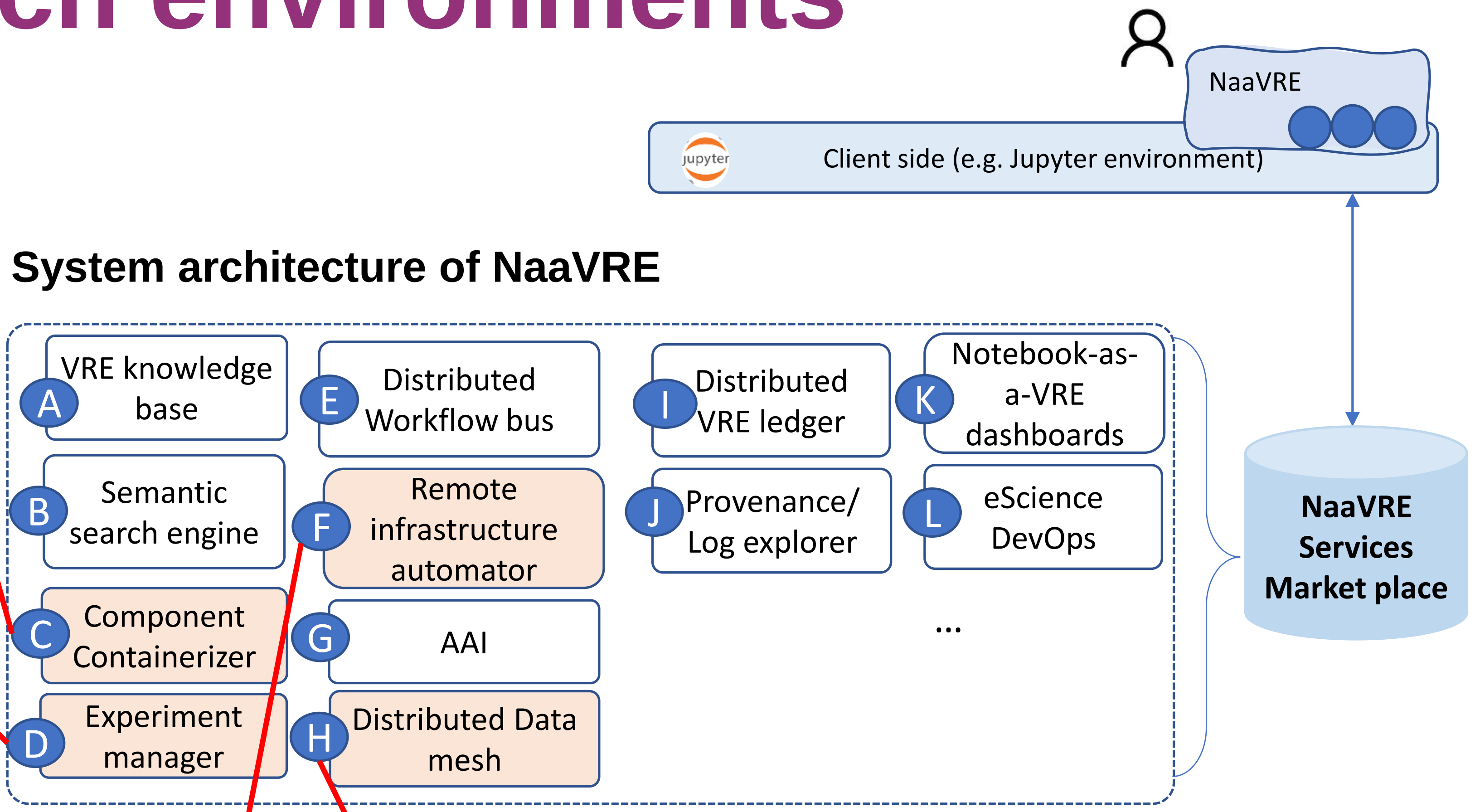
Zhao et al. 2022 *Software: Practice and Experience*

Virtual research environments

Component containerizer
- turns cells of python code into standardized RESTful API services and Docker containers to make them reusable (FAIR-CELLs)

Experiment manager
- provides a workflow of services with customized data flow and data input

System architecture of NaaVRE



Remote infrastructure automator
- automates the workflow execution on remote cloud environments (e.g. container deployment and workflow scheduling)

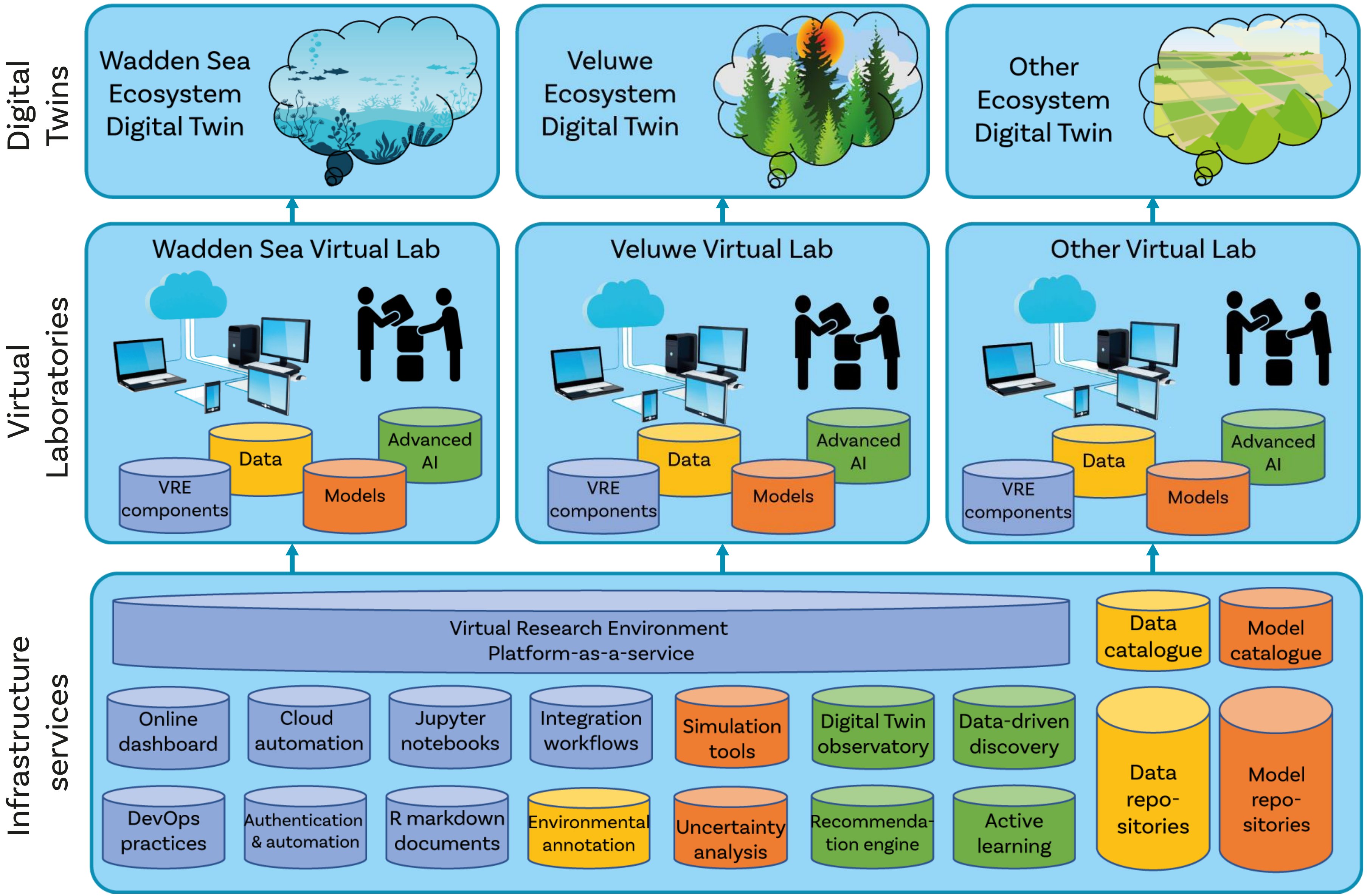
Distributed Data mesh
- supports data management in the workflow (centralized option = WebDAV, decentralized option = IPFS)

Zhao et al. 2022 *Software: Practice and Experience*

Digital Twins of ecosystems

Simulation platform for complex (eco)system analysis

Interactive, collaborative, virtual research environment



University of Amsterdam
Faculty of Science

LifeWatch ERIC

IBED
Institute for Biodiversity and Ecosystem Dynamics

Biodiversity data workflows for Digital Twins

W. Daniel Kissling[1,2]

[1]Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam

[2]Virtual Laboratory & Innovations Center (LW-VLIC), LifeWatch-ERIC

Digital Twin Conference, Wageningen, 14 December 2022

Thank you for your attention

University of Amsterdam

NWO
Netherlands Organisation for Scientific Research

netherlands eScience center

LifeWatch ERIC