Data & Science for Environmental health & SDG's

PD Dr. Angela Lausch, Peter Dietrich, Toralf Kirsten, Sonja Knapp, Josef Settele, Steffen Zacharias, Jan Bumberger
Angela.Lausch@ufz.de

HELMHOLTZ ZENTRUM FÜR UMWELTFORSCHUNG UFZ

Data Science for Environmental health - Challenges

- Landscapes, Environment, Processes & Functions are: Complex, Multidimensional, Multiscalar & mostly Non-linear

- No one monitoring approach, monitoring platform, model, space-time scale, tool or data alone are sufficient to explain the complexity of landscapes, processes or functions

- We have to look for necessary requirements – dealing of Complexity, Multidimensionality …,

Data Science – Challenge - Digitalization

Ecology and Evolution

Open Access

The PREDICTS database: a global database of how local terrestrial biodiversity responds to human impacts

Lawrence N. Hudson[1*], Tim Newbold[2,3*], Sara Contu[1], Samantha L. L. Hill[1,2], Igor Lysenko[4], Adriana De Palma[1,4], Helen R. P. Phillips[1,4], Rebecca A. Senior[2], Dominic J. Bennett[4], Hollie Booth[2,5], Argyrios

Global Change Biology

Global Change Biology (2011) 17, 2905–2935, doi: 10.1111/j.1365-2486.2011.02451.x

TRY – a global database of plant traits

J. KATTGE*, S. DÍAZ†, S. LAVOREL‡, I. C. PRENTICE§, P. LEADLEY¶, G. BÖNISCH*,

World database of protected areas (WDPA)

Movebank – For Animal Tracking data
www.movebank.org

Encyclopedia of life (EOL)

Contents lists available at ScienceDirect

Biological Conservation

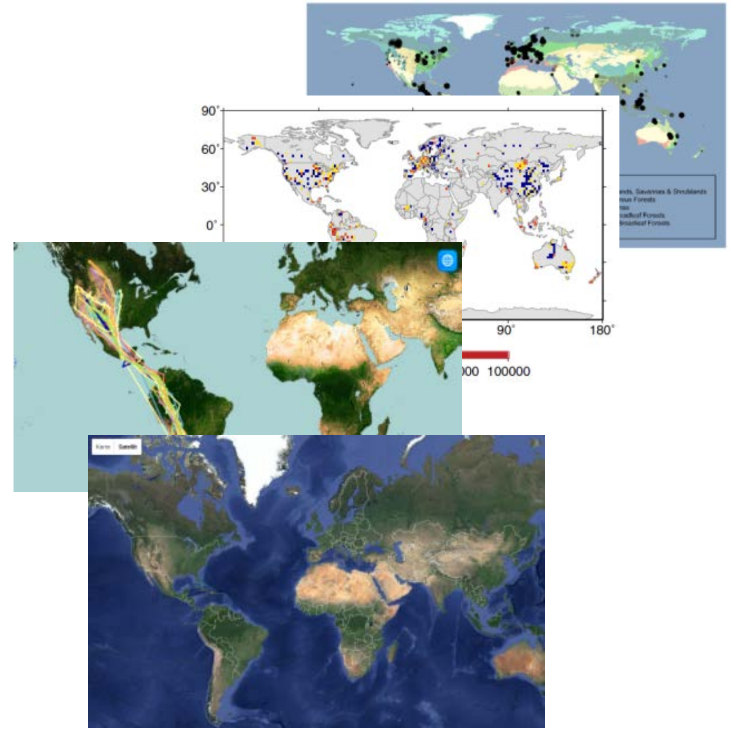
journal homepage: www.elsevier.com/locate/biocon

ELSEVIER

Perspective

Free and open-access satellite data are key to biodiversity conservation

W. Turner[a,*], C. Rondinini[b], N. Pettorelli[c], B. Mora[d], A.K. Leidner[a,e], Z. Szantoi[f], G. Buchanan[g],



Big Data

Free Data

Open Data

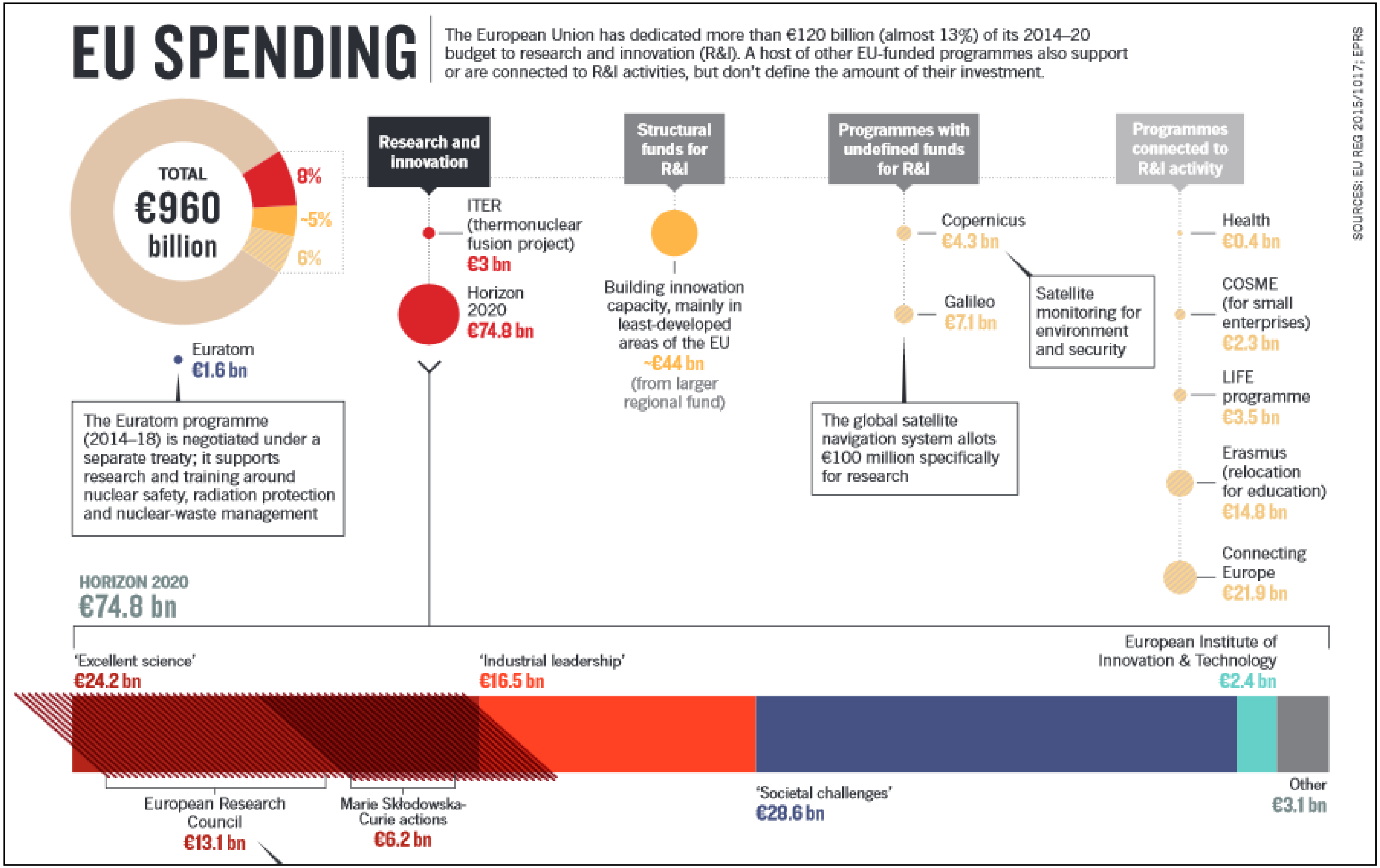
Complex Data

Data Science – Challenge - Digitalization

Roche, D.G., Lanfear, R., Binning, S.A., Haff, T.M., Schwanz, L.E., Cain, K.E., Kokko, H., Jennions, M.D., Kruuk, L.E.B., 2014. Troubleshooting Public Data Archiving: Suggestions to Increase Participation. PLoS Biol. 12. doi:10.1371/journal.pbio.1001779

Data Science – Challenge - Digitalization

Investitions/EU - Data-Generation 2014-2020 → 120B €



Abbott, A.; Butler, D.; Gibney, E.; Schiermeier, Q.; Van Noorden, R. Boon or burden: done for science? Nature 2016, 534, 307–309.

Data Science – Challenge - Digitalization

Although 90% of the world's data was generated over two years, around

→"50% of all research and experiment data (= US$28B/year) are not reproducible

→and over 80% of it never makes it to a trusted and sustainable repository" (Ayris et al., 2016)

Ayris, P.; Berthou, J.-Y.; Bruce, R.; Lindstaedt, S.; Monreale, A.; Mons, B.; M. Tochtermann, K.;Wilkinson, R. Realising the European Open Science Cloud; Europe
Brussels, Belgium, 2016.

Data Science – Requirements for Scientists

Received: 20 February 2018 | Accepted: 20 April 2018
DOI: 10.1111/2041-210X.13025

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

IMPROVING BIODIVERSITY MONITORING USING SATELLITE REMOTE SENSING

Understanding and assessing vegetation health by in situ species and remote-sensing approaches

Angela Lausch[1,2*] | Olaf Bastian[3] | Stefan Klotz[4] | Pedro J. Leitão[2,5] | András Jung[6,7] | Duccio Rocchini[8,9,10] | Michael E. Schaepman[11] | Andrew K. Skidmore[12,13] | Lutz Tischendorf[14] | Sonja Knapp[15]

[1]Department of Computational Landscape Ecology, Helmholtz Centre for Environmental Research—UFZ, Leipzig, Germany; [2]Geography Department, Humboldt University Berlin, Berlin, Germany; [3]OT Boxdorf, Moritzburg, Germany; [4]Department of Community Ecology, Helmholtz Centre for Environmental Research—UFZ, Halle, Germany; [5]Department Landscape Ecology and Environmental Systems Analysis, Technische Universität Braunschweig,

Lausch, A.; et al., 2018. Understanding and assessing vegetation health by in-situ species and remote sensing approaches. Methods in Ecology and Evolution, 00: 1–11. doi.org/10.1111/2041-210X.13025.

remote sensing MDPI

Review

Understanding Forest Health with Remote Sensing, Part III: Requirements for a Scalable Multi-Source Forest Health Monitoring Network Based on Data Science Approaches

Angela Lausch [1,2,*], Erik Borg [3], Jan Bumberger [4], Peter Dietrich [4,5], Marco Heurich [6,7], Andreas Huth [8], András Jung [9,10], Reinhard Klenke [11], Sonja Knapp [12], Hannes Mollenhauer [4], Hendrik Paasche [4], Heiko Paulheim [13], Marion Pause [14], Christian Schweitzer [15], Christiane Schmulius [16], Josef Settele [11,17], Andrew K. Skidmore [18,19], Martin Wegmann [20], Steffen Zacharias [4], Toralf Kirsten [21] and Michael E. Schaepman [22]

1 Department Computational Landscape Ecology, Helmholtz Centre for Environmental Research—UFZ, Permoserstr. 15, D-04318 Leipzig, Germany

Lausch, A. et al., 2018. Understanding Forest Health with Remote Sensing, Part III: Requirements for a Scalable Multi-Source Forest Health Monitoring Network Based on Data Science Approaches. Remote Sensing, 10, 1120; doi:10.3390/rs10071120.

Data Science – Requirements for Scientists

nature ecology & evolution

PERSPECTIVE

https://doi.org/10.1038/s41559-018-0667-3

OPEN

Towards global data products of Essential Biodiversity Variables on species traits

W. Daniel Kissling[1*], Ramona Walls[2], Anne Bowser[3], Matthew O. Jones[4], Jens Kattge[5,6], Donat Agosti[7], Josep Amengual[8], Alberto Basset[9], Peter M. van Bodegom[10], Johannes H. C. Cornelissen[11], Ellen G. Denny[12], Salud Deudero[13], Willi Egloff[7], Sarah C. Elmendorf[14,15], Enrique Alonso García[16], Katherine D. Jones[14], Owen R. Jones[17], Sandra Lavorel[18], Dan Lear[19], Laetitia M. Navarro[6,20], Samraat Pawar[21], Rebecca Pirzl[22], Nadja Rüger[6,23], Sofia Sal[21], Roberto Salguero-Gómez[24,25,26,27], Dmitry Schigel[28], Katja-Sabine Schulz[29], Andrew Skidmore[30,31] and Robert P. Guralnick[32]

Kissling, W.D., et al., 2018. Towards global data products of Essential Biodiversity Variables (EBVs) on species traits. Nat. Ecol. Evol. doi:10.1038/s41559-018-0667-3

Data Science – Requirements (selection)

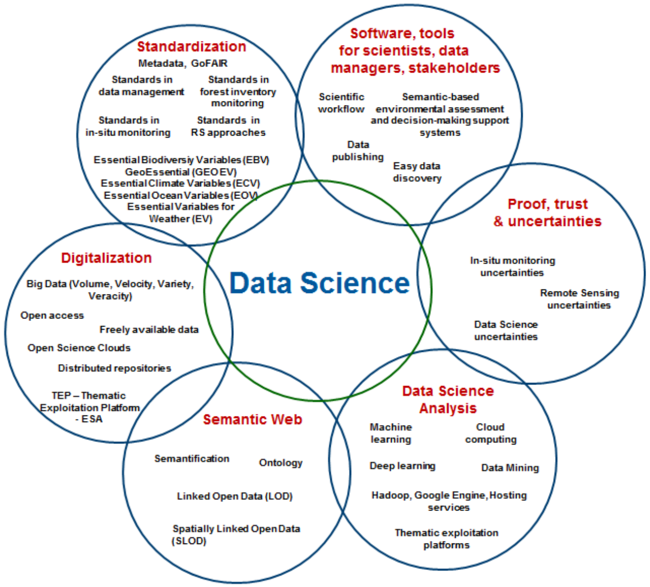
Good Indicators for environmental health, changes, stress & disturbances, SDG's

Digitalization

(Big Data (Volume, Velocity, Variety, Veracity), Open Access, Freely available data, Open Science Clouds, Distributed repositories, TEP – Thematic Exploitation Platform – ESA)

Standardization

(Metadata, GoFAIR, Concept of Essential Variables – EV Essential Biodiversity Variables)



Lausch, A. et al., 2018.. Remote Sensing

Semantification

(Semantic Web/Web 4.0, Ontology; Linked Open Data –LOD)

Data Science Analysis

(Machine Learning, Deep learning, Cloud Computing, Data Mining, Hadoop, Google Engine, Hosting services)

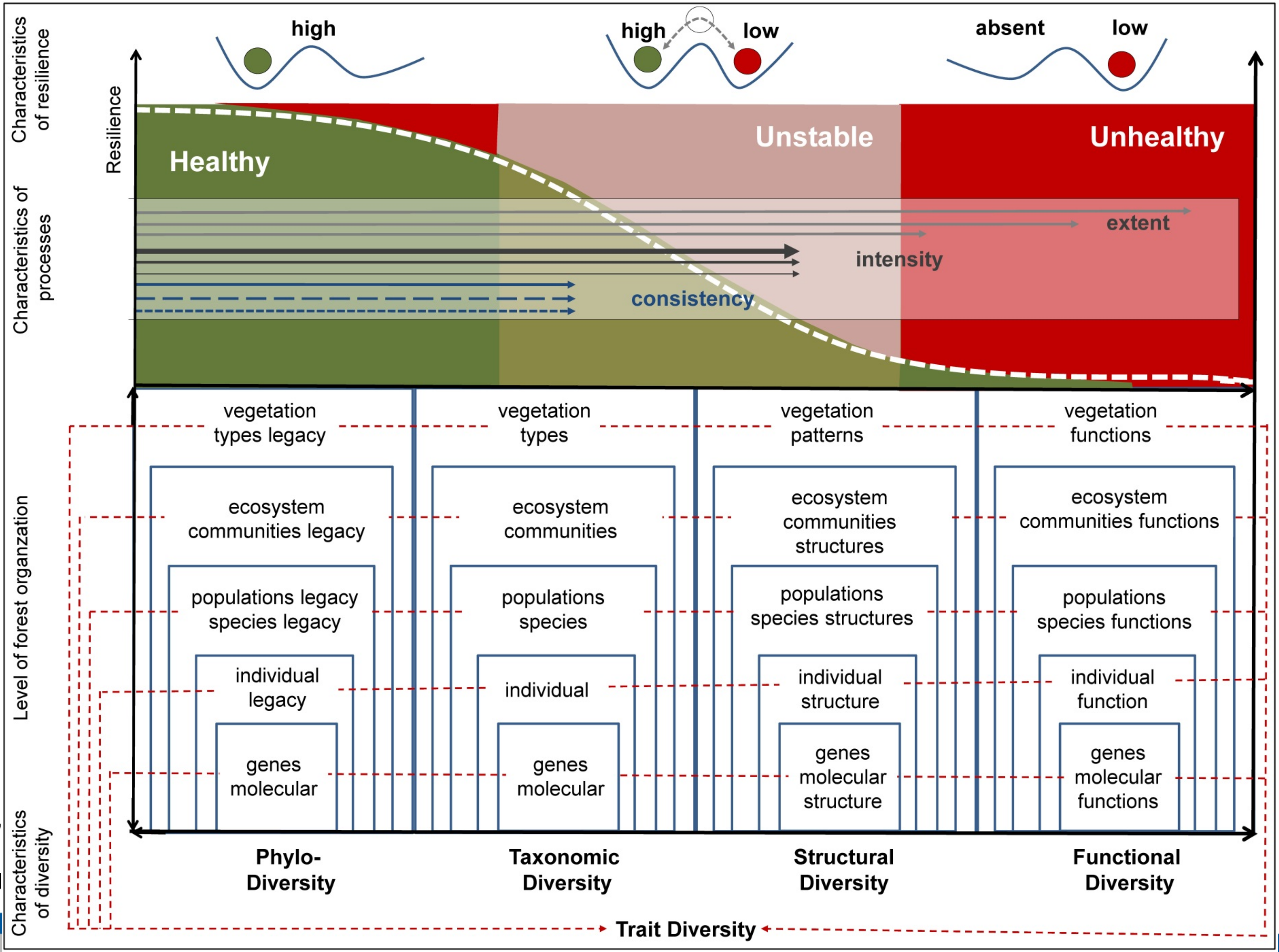
Proof, trust & uncertainties

(In-situ monitoring, Remote Sensing & Data Science uncertainties)

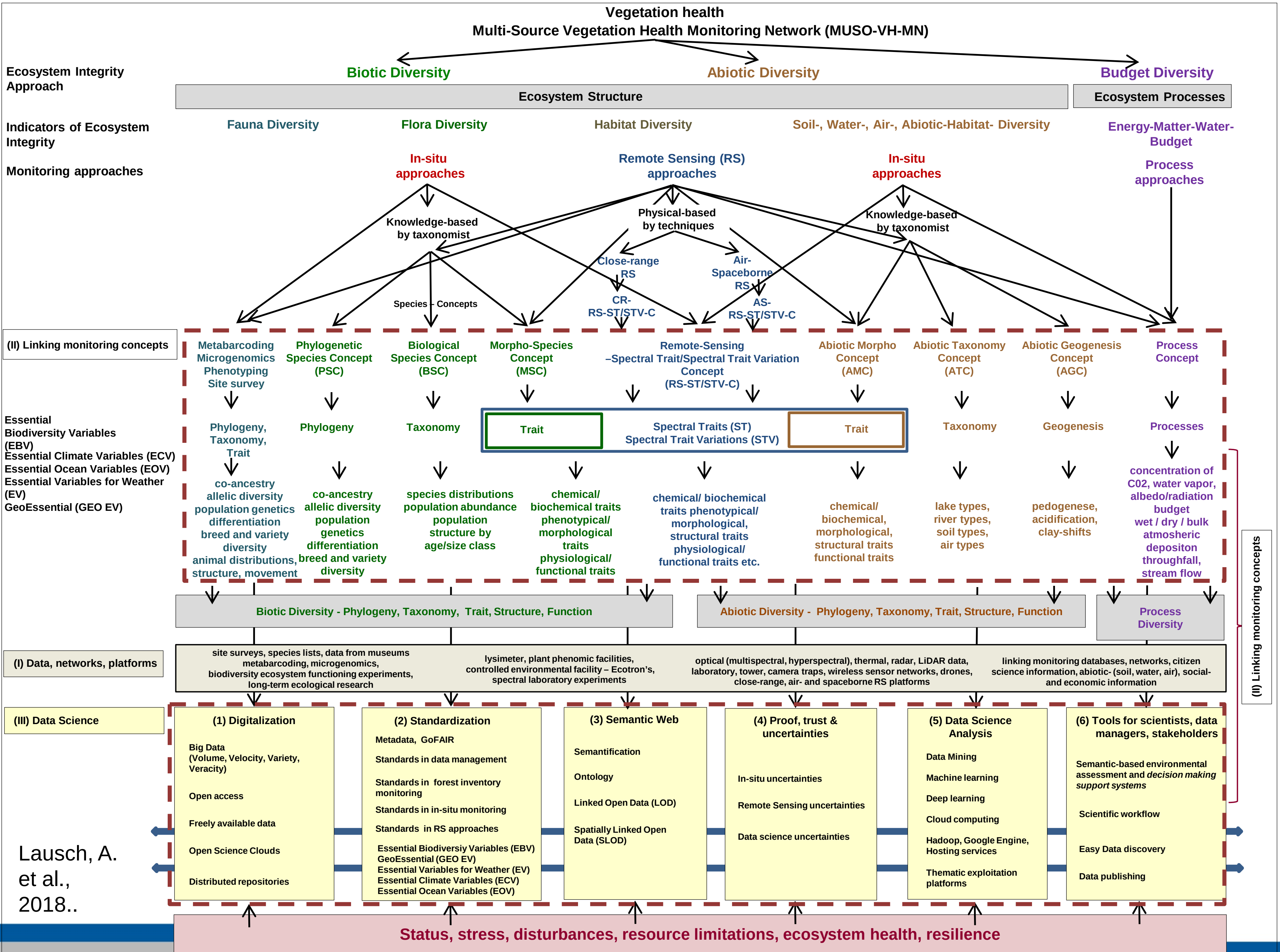
Easy software, tools for data manager, stakolders

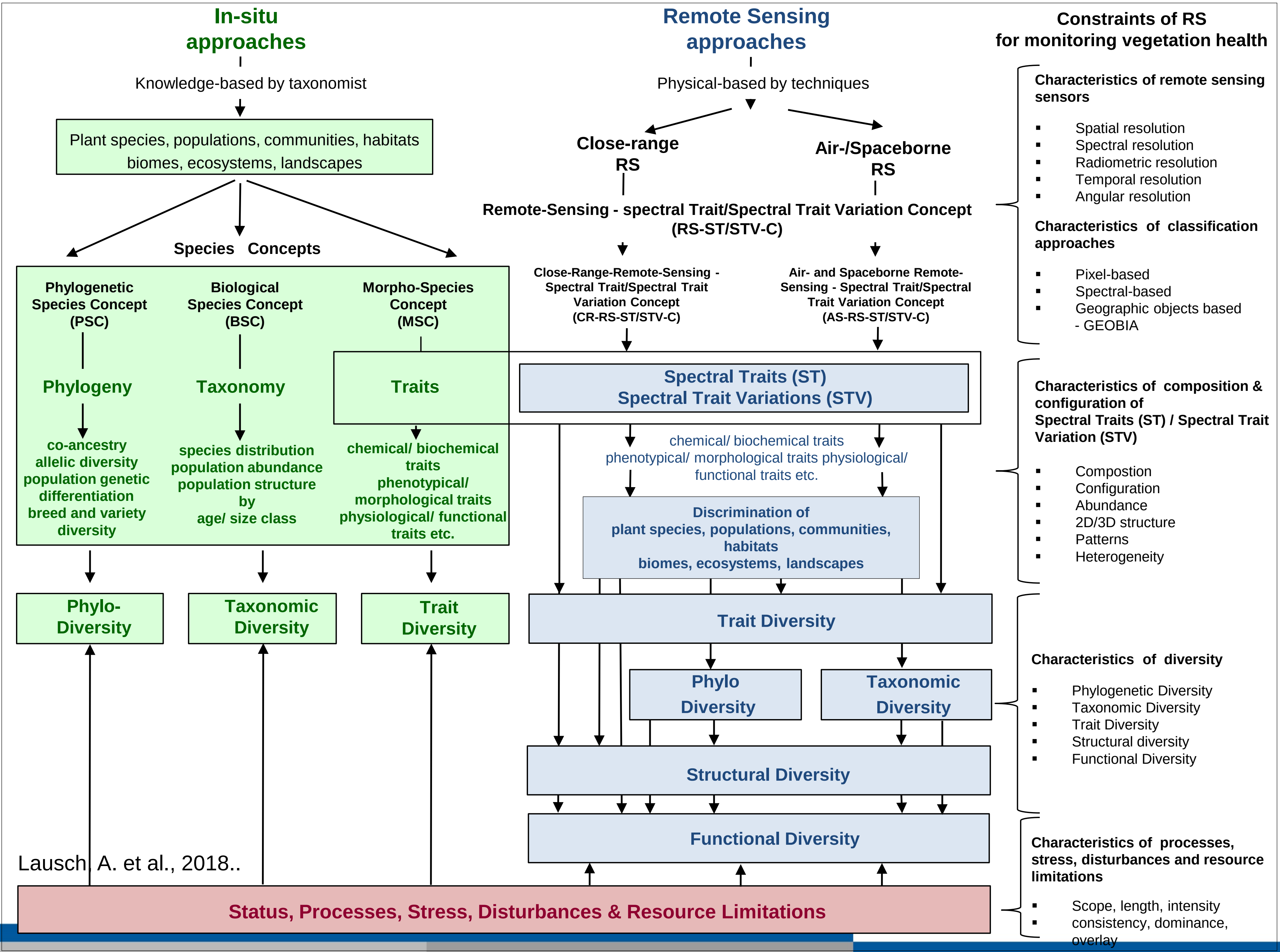
HELMHOLTZ ZENTRUM FÜR UMWELTFORSCHUNG UFZ

Data Science – Requirement
Indicators of environmental stress



Lausch, A. et al., 2018..
Remote Sensing

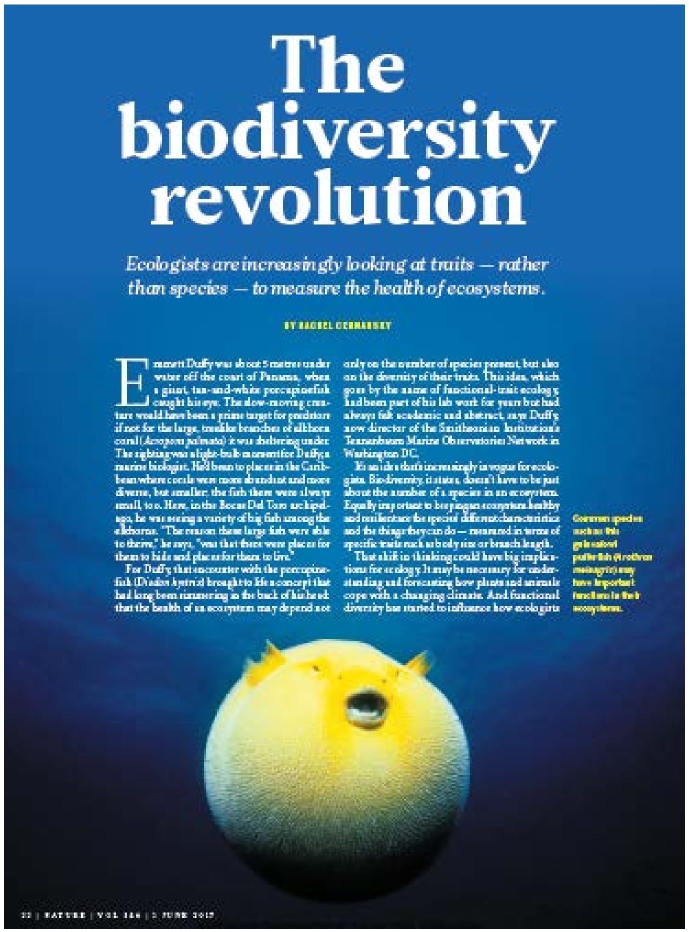

Vegetation health
Multi-Source Vegetation Health Monitoring Network (MUSO-VH-MN)
Ecosystem Integrity Approach
Biotic Diversity
Abiotic Diversity
Budget Diversity
Ecosystem Structure
Ecosystem Processes
Indicators of Ecosystem Integrity
Fauna Diversity
Flora Diversity
Habitat Diversity
Soil-, Water-, Air-, Abiotic-Habitat- Diversity
Energy-Matter-Water-Budget
Monitoring approaches
In-situ approaches
Remote Sensing (RS) approaches
In-situ approaches
Process approaches
Knowledge-based by taxonomist
Physical-based by techniques
Knowledge-based by taxonomist
Close-range RS
Air-Spaceborne RS
Species - Concepts
CR-RS-ST/STV-C
AS-RS-ST/STV-C
(II) Linking monitoring concepts
Metabarcoding Microgenomics Phenotyping Site survey
Phylogenetic Species Concept (PSC)
Biological Species Concept (BSC)
Morpho-Species Concept (MSC)
Remote-Sensing –Spectral Trait/Spectral Trait Variation Concept (RS-ST/STV-C)
Abiotic Morpho Concept (AMC)
Abiotic Taxonomy Concept (ATC)
Abiotic Geogenesis Concept (AGC)
Process Concept
Essential Biodiversity Variables (EBV)
Essential Climate Variables (ECV)
Essential Ocean Variables (EOV)
Essential Variables for Weather (EV)
GeoEssential (GEO EV)
Phylogeny, Taxonomy, Trait
Phylogeny
Taxonomy
Trait
Spectral Traits (ST) Spectral Trait Variations (STV)
Trait
Taxonomy
Geogenesis
Processes
co-ancestry allelic diversity population genetics differentiation breed and variety diversity animal distributions, structure, movement
co-ancestry allelic diversity population genetics differentiation breed and variety diversity
species distributions population abundance population structure by age/size class
chemical/ biochemical traits phenotypical/ morphological traits physiological/ functional traits
chemical/ biochemical traits phenotypical/ morphological, structural traits physiological/ functional traits etc.
chemical/ biochemical, morphological, structural traits functional traits
lake types, river types, soil types, air types
pedogenese, acidification, clay-shifts
concentration of C02, water vapor, albedo/radiation budget wet / dry / bulk atmospheric depositon throughfall, stream flow
Biotic Diversity - Phylogeny, Taxonomy, Trait, Structure, Function
Abiotic Diversity - Phylogeny, Taxonomy, Trait, Structure, Function
Process Diversity
(II) Linking monitoring concepts
(I) Data, networks, platforms
site surveys, species lists, data from museums metabarcoding, microgenomics, biodiversity ecosystem functioning experiments, long-term ecological research
lysimeter, plant phenomic facilities, controlled environmental facility – Ecotron's, spectral laboratory experiments
optical (multispectral, hyperspectral), thermal, radar, LiDAR data, laboratory, tower, camera traps, wireless sensor networks, drones, close-range, air- and spaceborne RS platforms
linking monitoring databases, networks, citizen science information, abiotic- (soil, water, air), social- and economic information
(III) Data Science
(1) Digitalization
Big Data (Volume, Velocity, Variety, Veracity)
Open access
Freely available data
Open Science Clouds
Distributed repositories
(2) Standardization
Metadata, GoFAIR
Standards in data management
Standards in forest inventory monitoring
Standards in in-situ monitoring
Standards in RS approaches
Essential Biodiversiy Variables (EBV) GeoEssential (GEO EV) Essential Variables for Weather (EV) Essential Climate Variables (ECV) Essential Ocean Variables (EOV)
(3) Semantic Web
Semantification
Ontology
Linked Open Data (LOD)
Spatially Linked Open Data (SLOD)
(4) Proof, trust & uncertainties
In-situ uncertainties
Remote Sensing uncertainties
Data science uncertainties
(5) Data Science Analysis
Data Mining
Machine learning
Deep learning
Cloud computing
Hadoop, Google Engine, Hosting services
Thematic exploitation platforms
(6) Tools for scientists, data managers, stakeholders
Semantic-based environmental assessment and decision making support systems
Scientific workflow
Easy Data discovery
Data publishing
Lausch, A. et al., 2018..
Status, stress, disturbances, resource limitations, ecosystem health, resilience



In-situ approaches
Knowledge-based by taxonomist
Plant species, populations, communities, habitats biomes, ecosystems, landscapes
Species Concepts
Phylogenetic Species Concept (PSC)
Biological Species Concept (BSC)
Morpho-Species Concept (MSC)
Phylogeny
Taxonomy
Traits
co-ancestry allelic diversity population genetic differentiation breed and variety diversity
species distribution population abundance population structure by age/ size class
chemical/ biochemical traits phenotypical/ morphological traits physiological/ functional traits etc.
Phylo-Diversity
Taxonomic Diversity
Trait Diversity
Remote Sensing approaches
Physical-based by techniques
Close-range RS
Air-/Spaceborne RS
Remote-Sensing - spectral Trait/Spectral Trait Variation Concept (RS-ST/STV-C)
Close-Range-Remote-Sensing - Spectral Trait/Spectral Trait Variation Concept (CR-RS-ST/STV-C)
Air- and Spaceborne Remote-Sensing - Spectral Trait/Spectral Trait Variation Concept (AS-RS-ST/STV-C)
Spectral Traits (ST)
Spectral Trait Variations (STV)
chemical/ biochemical traits phenotypical/ morphological traits physiological/ functional traits etc.
Discrimination of plant species, populations, communities, habitats biomes, ecosystems, landscapes
Trait Diversity
Phylo Diversity
Taxonomic Diversity
Structural Diversity
Functional Diversity
Status, Processes, Stress, Disturbances & Resource Limitations
Constraints of RS for monitoring vegetation health
Characteristics of remote sensing sensors
Spatial resolution
Spectral resolution
Radiometric resolution
Temporal resolution
Angular resolution
Characteristics of classification approaches
Pixel-based
Spectral-based
Geographic objects based - GEOBIA
Characteristics of composition & configuration of Spectral Traits (ST) / Spectral Trait Variation (STV)
Compostion
Configuration
Abundance
2D/3D structure
Patterns
Heterogeneity
Characteristics of diversity
Phylogenetic Diversity
Taxonomic Diversity
Trait Diversity
Structural diversity
Functional Diversity
Characteristics of processes, stress, disturbances and resource limitations
Scope, length, intensity
consistency, dominance, overlay
Lausch, A. et al., 2018..

Data Science – Requirement - Indicators of stress

“Ecologists are increasingly looking **at traits** - **rather than species** - to measure the health of ecosystems”


The biodiversity revolution

Ecologists are increasingly looking at traits — rather than species — to measure the health of ecosystems.

Traits

↓

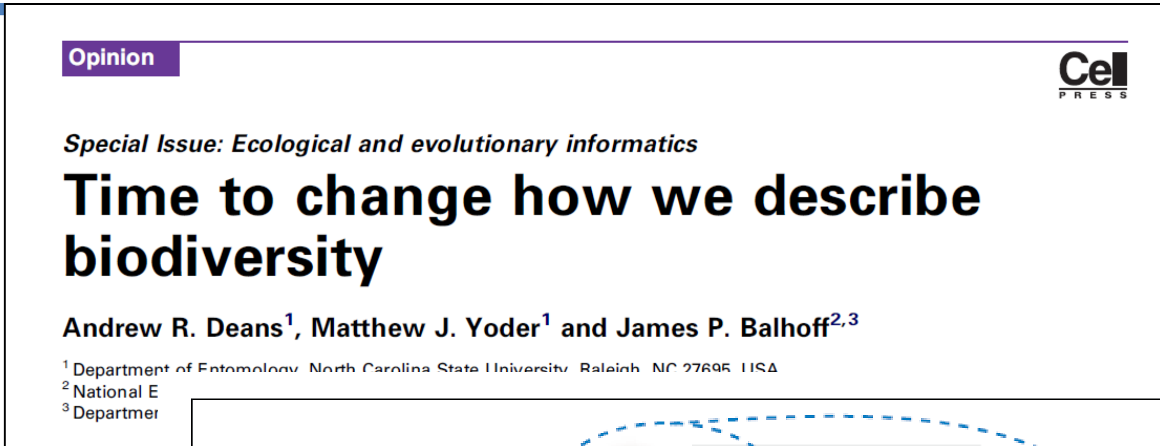
Indicators of
Stress
Disturbances
Ressource limitation

↓

Trait-Variations

Cernansky, R. Biodiversity moves beyond counting species. *Nature* 2017, *546*, 22–24

Data Science – Requirement – Traits / Phenotyping - Animals


Opinion

Cell PRESS

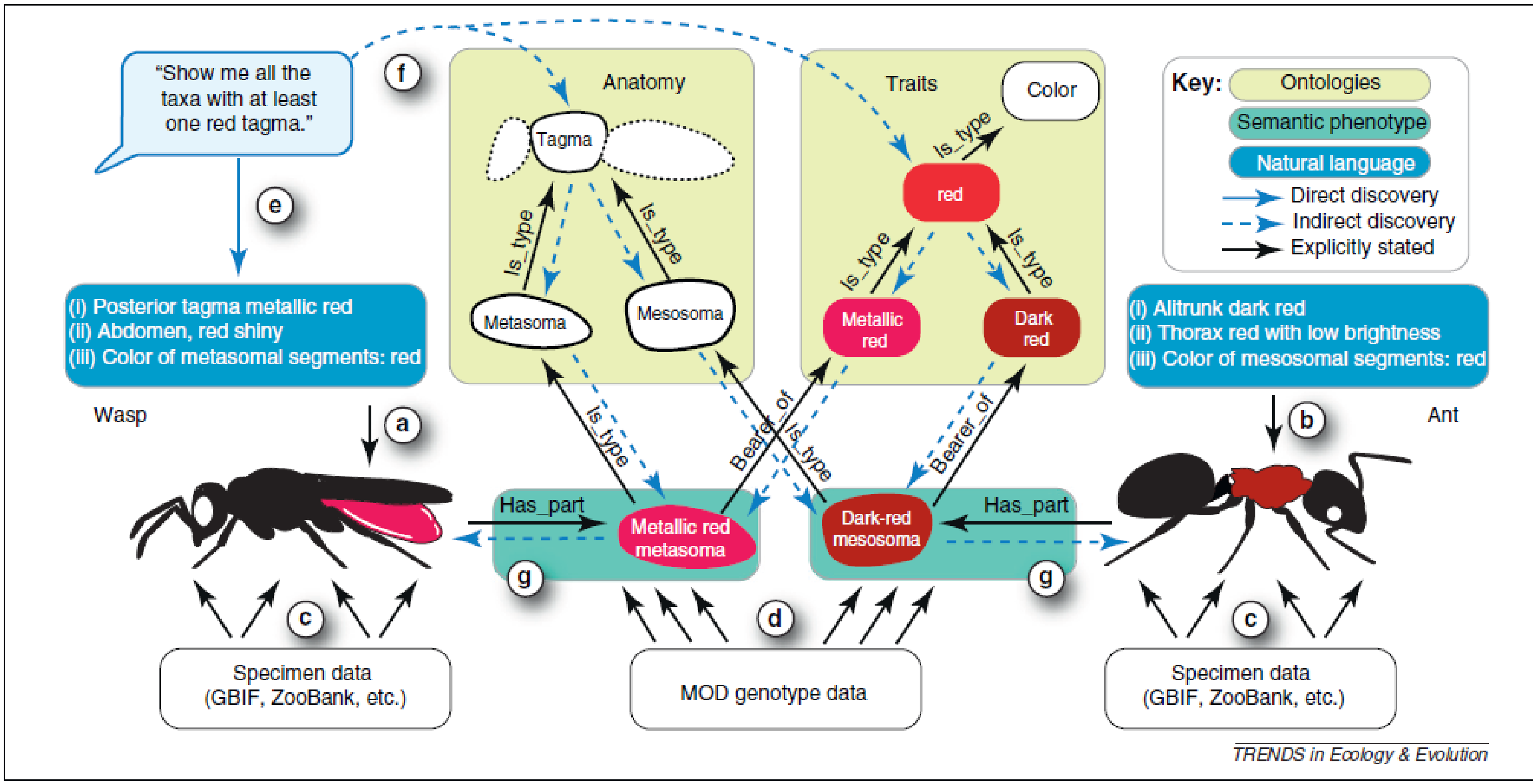
Special Issue: Ecological and evolutionary informatics

Time to change how we describe biodiversity

Andrew R. Deans[1], Matthew J. Yoder[1] and James P. Balhoff[2,3]

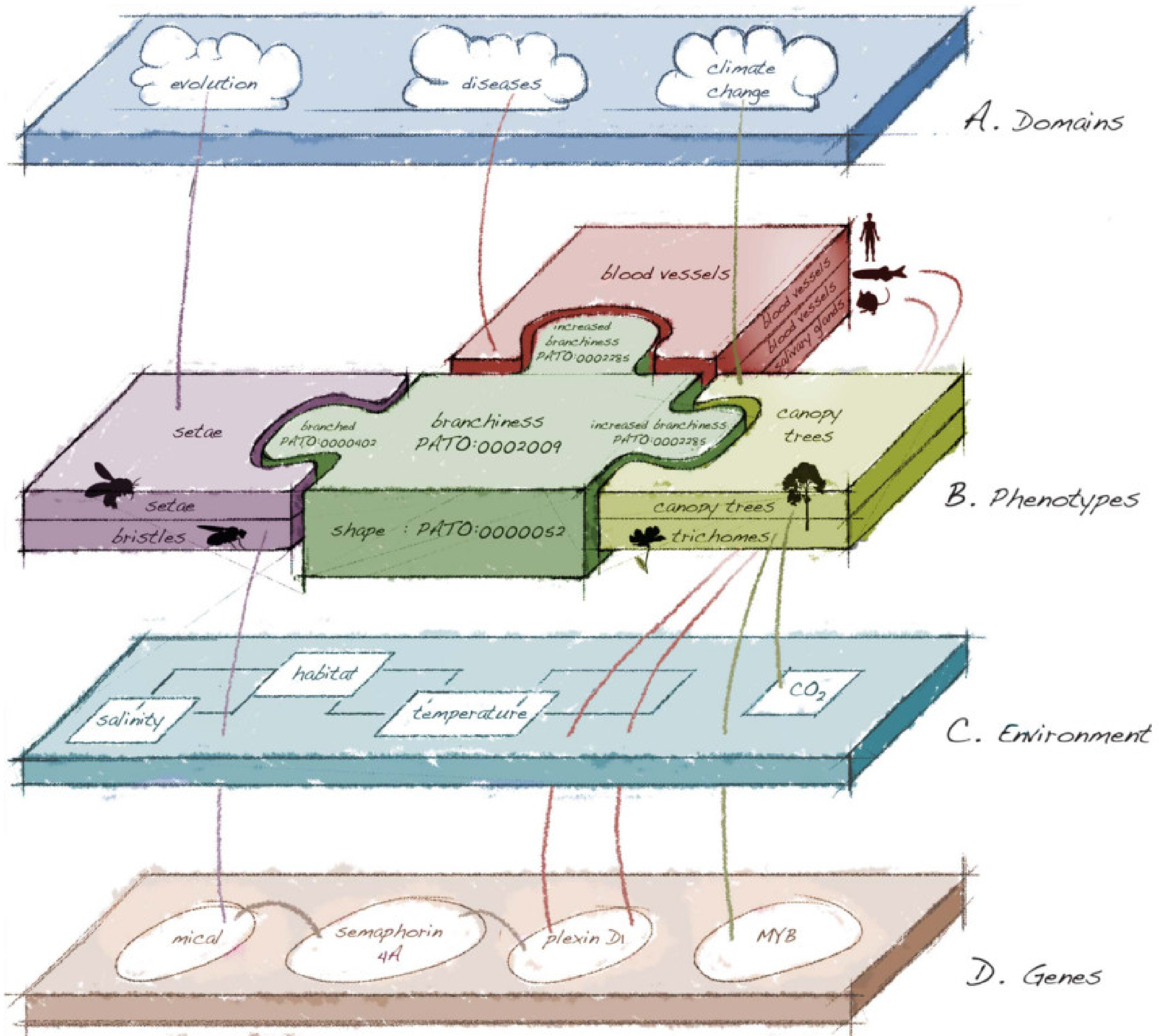
[1] Department of Entomology, North Carolina State University, Raleigh, NC 27695, USA
[2] National E
[3] Departmer

→ Semantic Web/ Linked Open Data



Deans, A.R., Yoder, M.J., Balhoff, J.P., 2012. Time to change how we describe biodiversity. Trends Ecol. Evol. 27, 78–84.

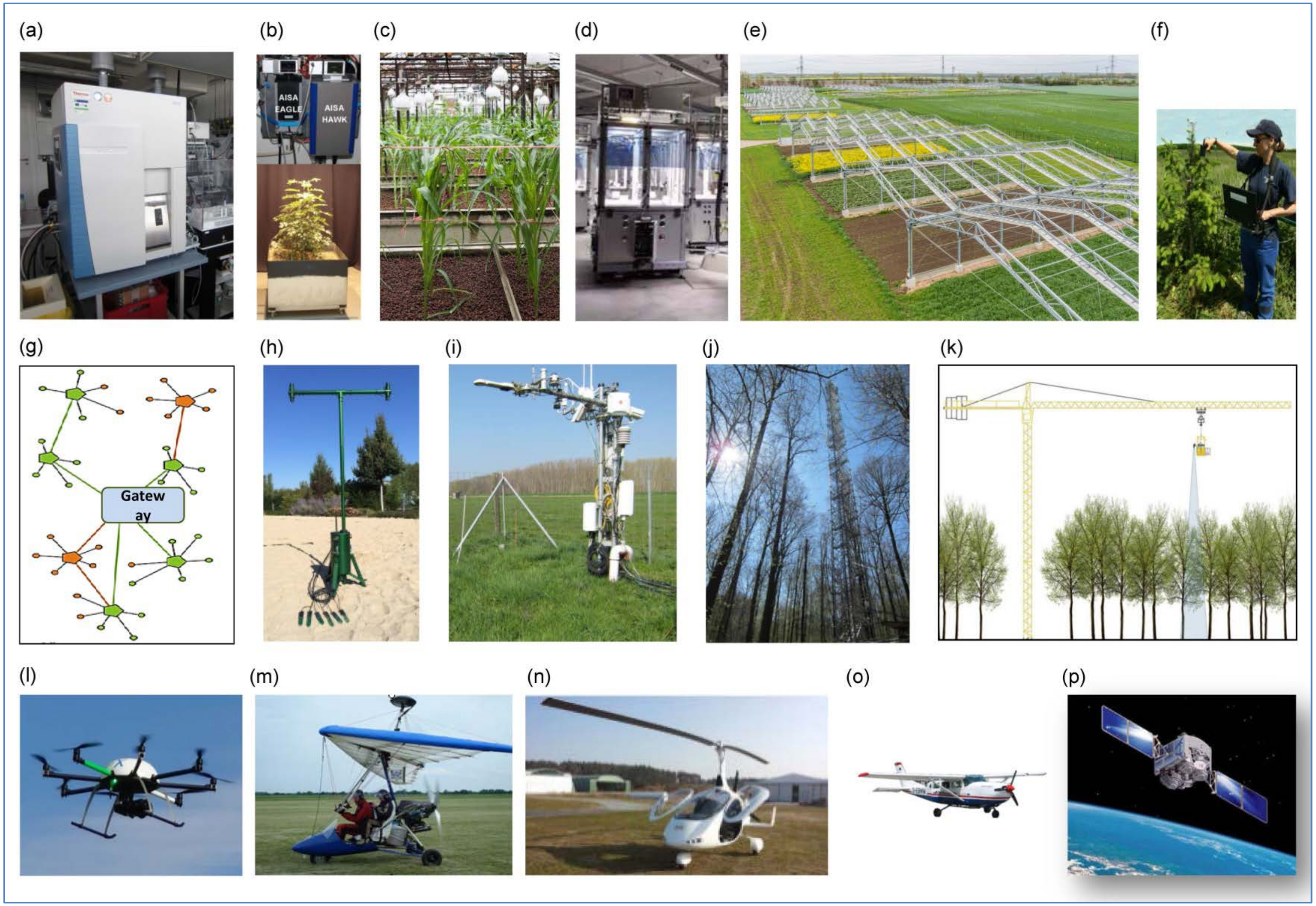
Data Science – Requirement Traits / Phenotyping
Indicators for environmental stress



on different scales

Deans et al., 2015. Finding Our Way through Phenotypes. *PLoS Biol.* 1284 *13*.

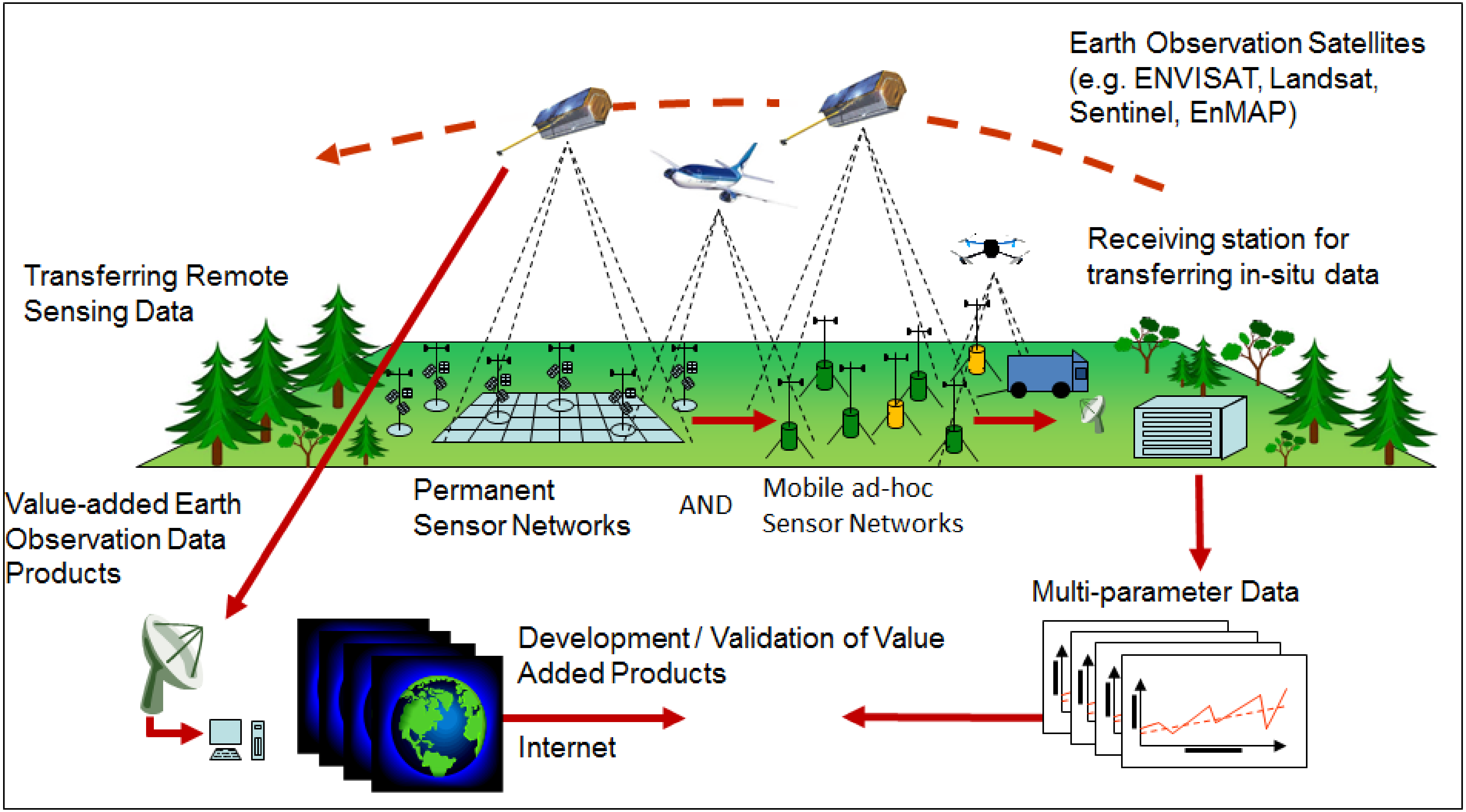
Data Science – Requirement – Remote Sensing



Different Platforms

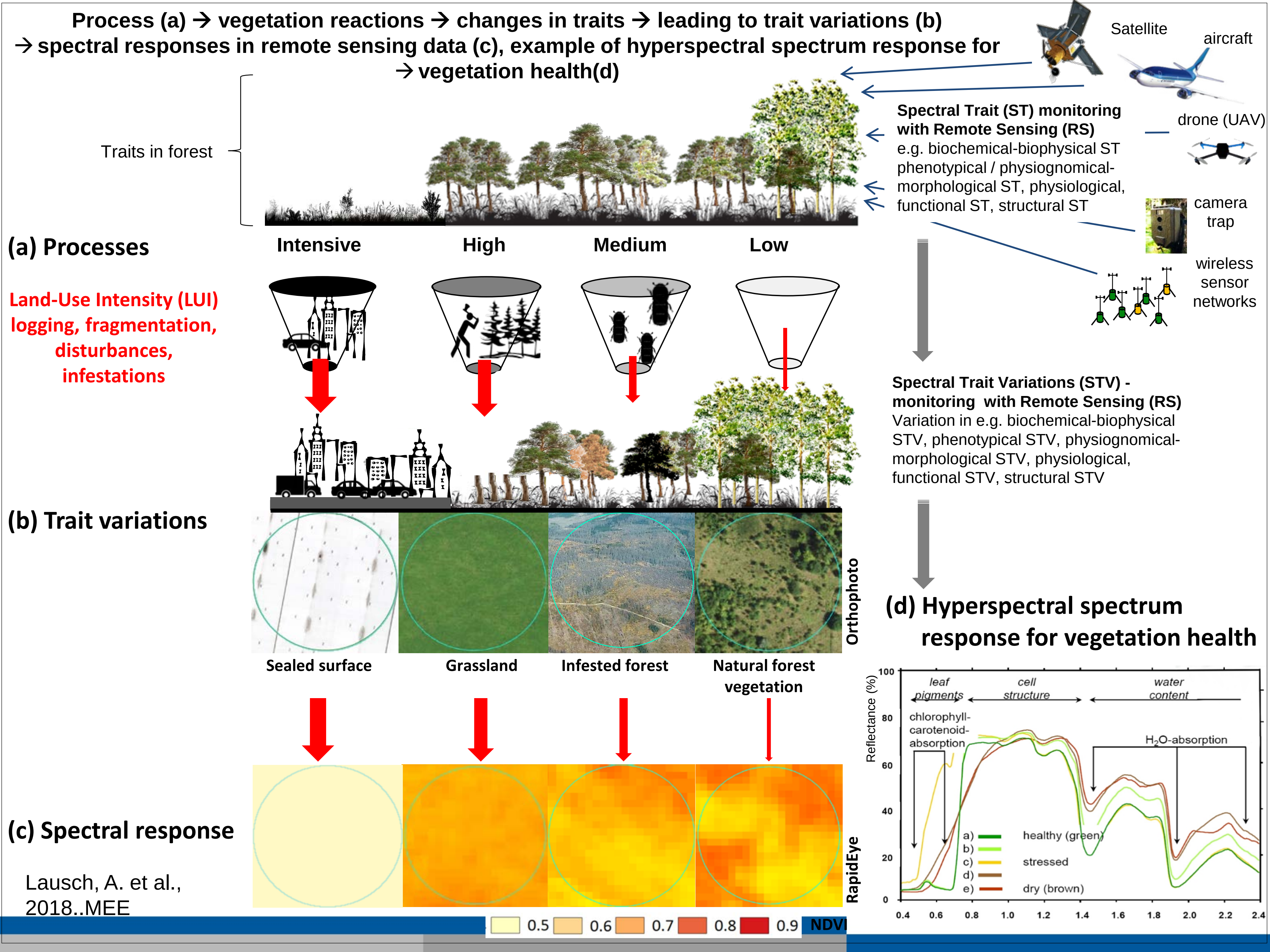
Lausch et al.,. A range of Earth Observation techniques for assessing plant diversity Jeannine Cavender-Bares, John Gamon, Philip Townsend (eds): The nature of biodiversity: prospects for remote detection of genetic, phylogenetic, functional and ecosystem components and importance in managing Planet, Jeannine Cavender-Bares, John Gamon, Philip Townsend, Springer, 2018/2019 (in press)

Data Science – Requirement – Coupling RS Platforms



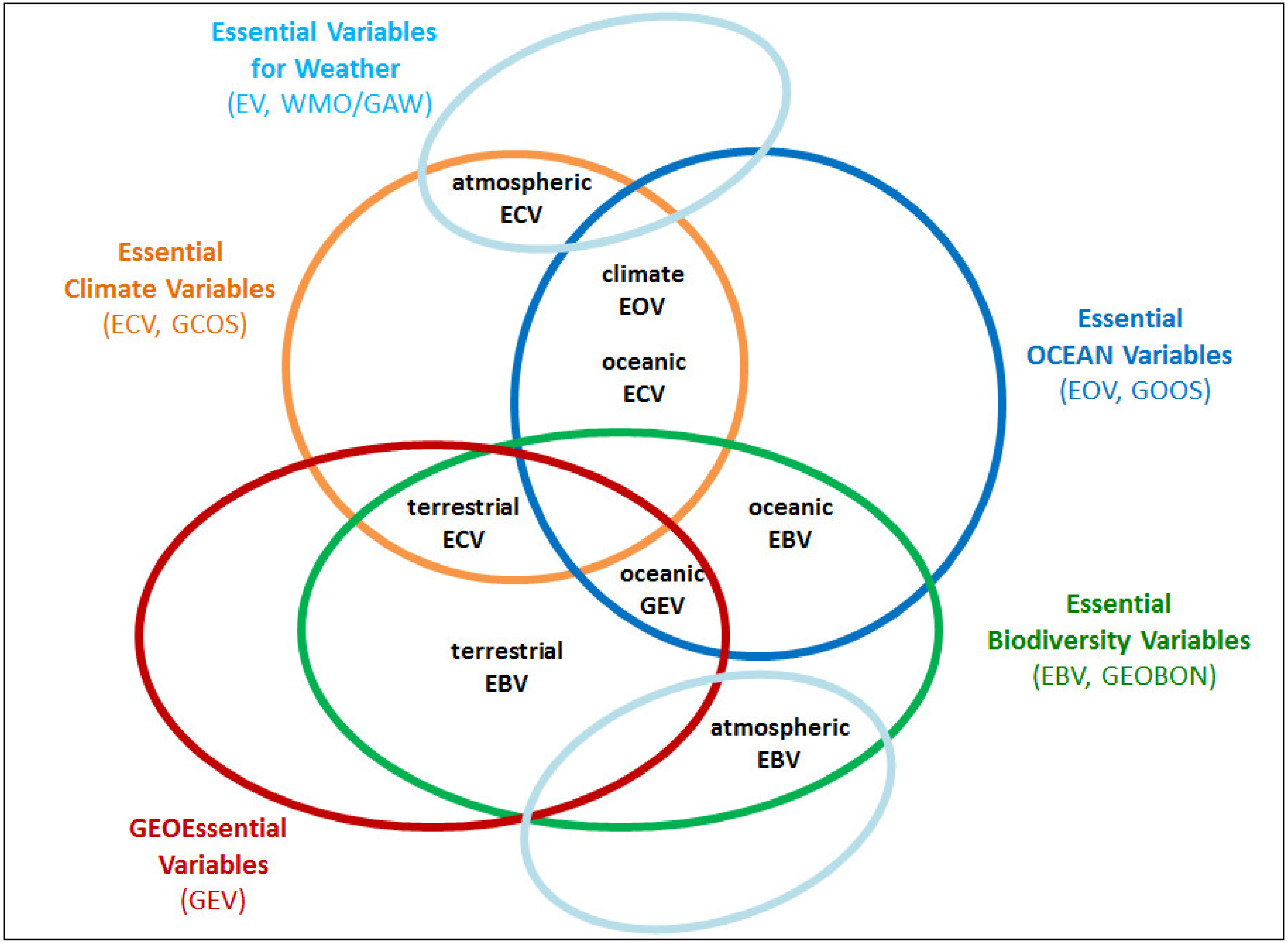
Lausch, A. et al., 2018. Understanding Forest Health with Remote Sensing, Part III: Requirements for a Scalable Multi-Source Forest Health Monitoring Network Based on Data Science Approaches. Remote Sensing, 10, 1120

Process (a) → vegetation reactions → changes in traits → leading to trait variations (b) → spectral responses in remote sensing data (c), example of hyperspectral spectrum response for → vegetation health(d)



Lausch, A. et al., 2018..MEE

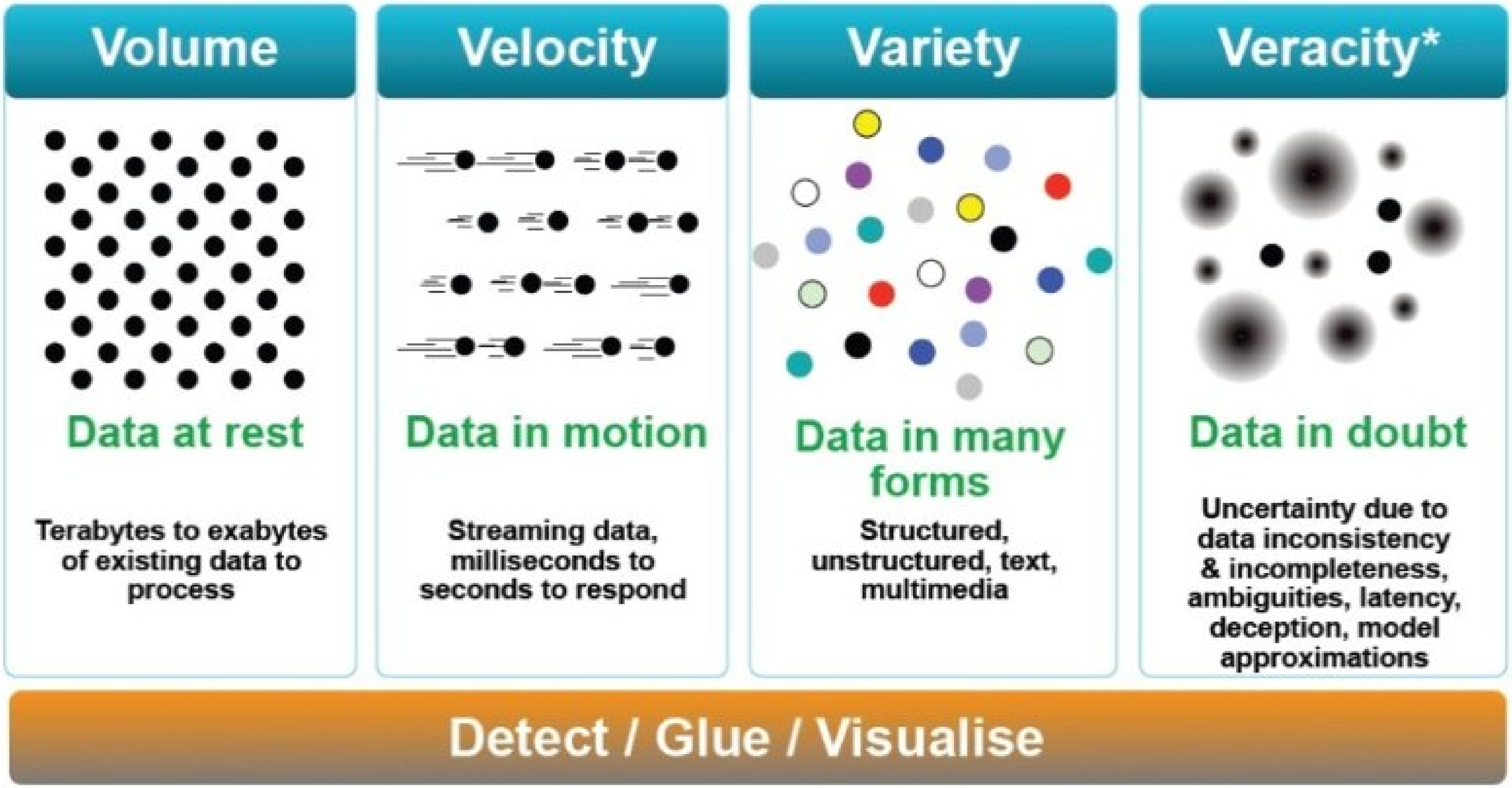
Data Science – Standardization in Monitoring Approaches



Lausch, A. et al., 2018. Understanding Forest Health with Remote Sensing, Part III: Requirements for a Scalable Multi-Source Forest Health Monitoring Network Based on Data Science Approaches. Remote Sensing, 10, 1120

Data Science – Requirement – Big Data

Characteristics of Big Data – 4 Vs



Data Science – Requirement – Metadata/Data - FAIR

FAIR – Guiding principles for dealing scientific data

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Research data

» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*#

Received: 10 December 2015

Accepted: 12 February 2016

Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Wilkinson, M.D., Dumontier, M., Aalversberg, I.J., Appleton, G., Axton, M., 2016. Comment : The FAIR Guiding Principles for scientific data management and stewardship. Nat. Commun. 3:160018, 1–9. doi:10.1038/sdata.2016.18

Data Science – Requirement – Metadata/Data - FAIR

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES
» Research data
» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*#

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Findable

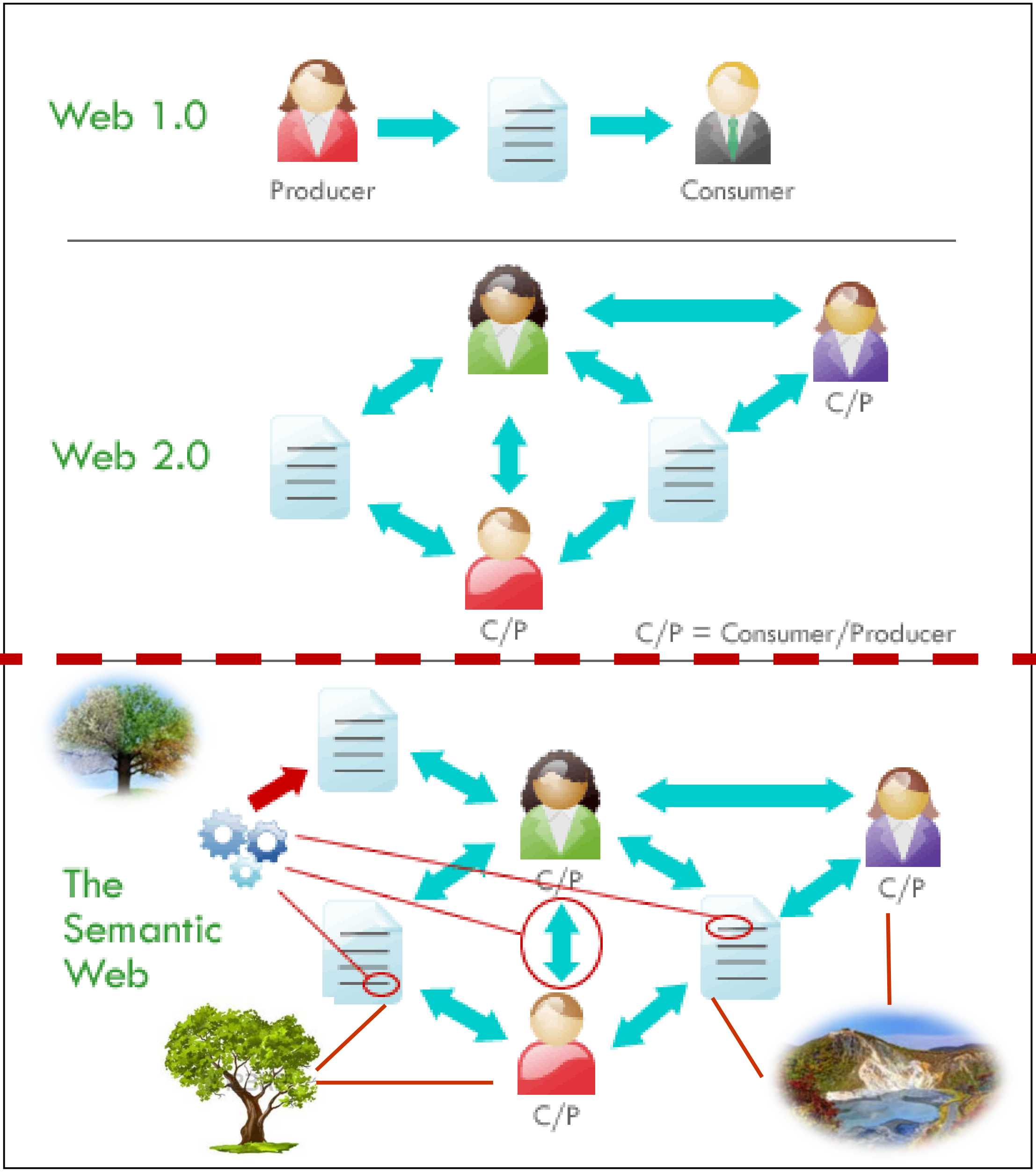
Accessible

Interoperable

Reusable

Wilkinson, M.D., Dumontier, M., Aalversberg, I.J., Appleton, G., Axton, M., 2016. Comment : The FAIR Guiding Principles for scientific data management and stewardship. Nat. Commun. 3:160018, 1–9.

Data Science – Requirement - Semantification



Semantic Web / Linked Open Data

Handling:

- **Complex-Data**

HELMHOLTZ ZENTRUM FÜR UMWELTFORSCHUNG UFZ

Approach for coupling complex data

– Semantic/Semantic Web – Linked Open Data

A Semantic Network to describe the relationships of Oscars‘s family members woth one others

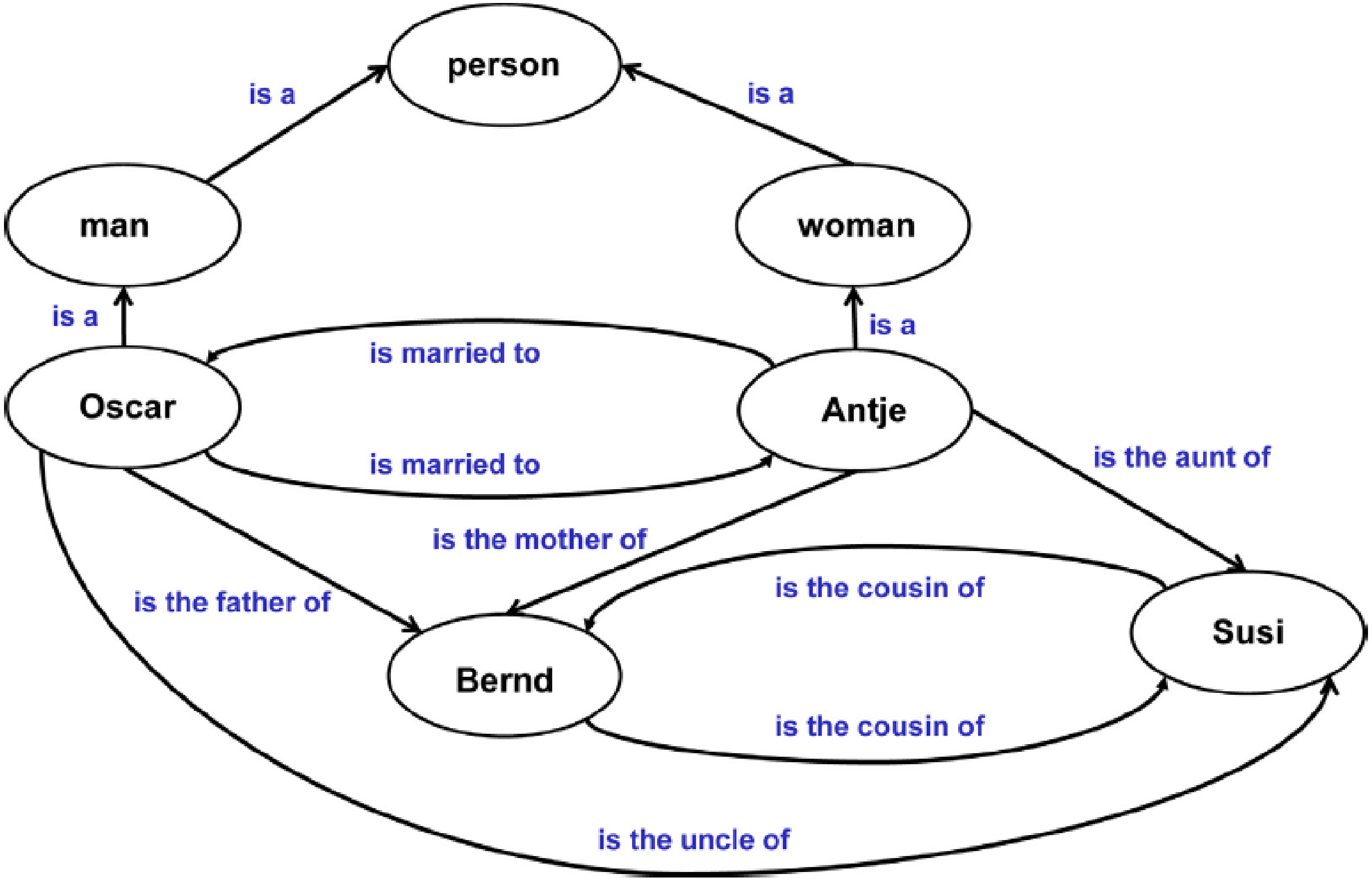


Fig. 7. Semantic network to describe the relationships of Oscar's family members with one another.

Lausch, Schmidt, Tischendorf, 2015. Data mining and linked open data – new perspectives for data analysis in environmental research. Ecological Modelling, 295, 5–17

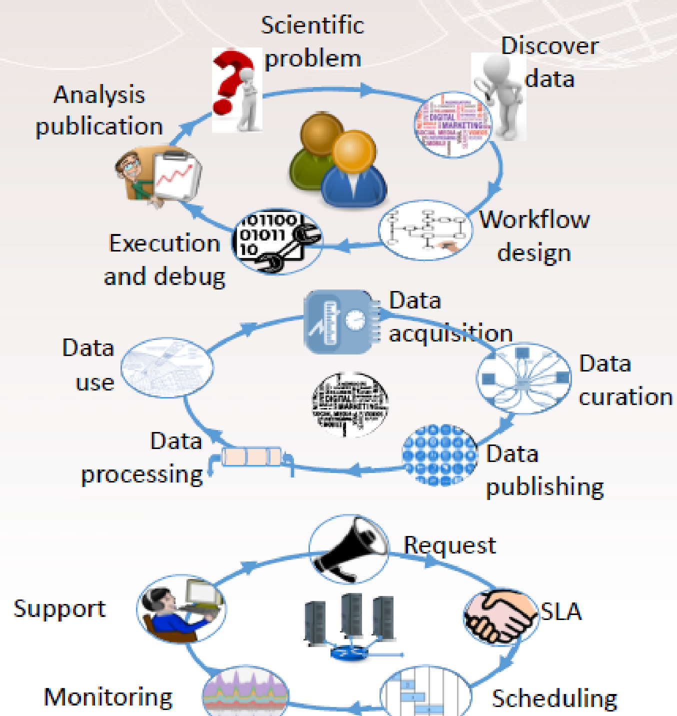
Data Science – Requirements – Data & Information Management

ENVRI plus – Environmental Research Infrastructures Providing Shared Solutions for Science and Society

http://www.envriplus.eu/

Research support environments

- **Need to** support user centered research activities
- **Need to** manage data in its lifecycle
- **Need to** manage infrastructure resources, e.g., computing, storage and networks



Zhiming Zhao, 2018, International Summer School “Data Management in Environmental and Earth, Science Infrastructures: theory and practice” *Dates, 9th July – 13th July 2018,* Lecce, Italy

Data & Science for Environmental health & SDG's

Thank you very much for your attention !

PD Dr. Angela Lausch, Peter Dietrich, Toralf Kirsten, Sonja Knapp, Josef Settele, Steffen Zacharias, Jan Bumberger
Angela.Lausch@ufz.de

HELMHOLTZ
ZENTRUM FÜR UMWELTFORSCHUNG
UFZ