From Observational Data to Information IG (OD2I IG)

The OD2I Team

tinyurl.com/y74p56tb

Tour de Table
(time permitted)

OD2I IG

- Primary data are interpreted for their meaning in determinate contexts
- Contexts relevant to science, industry, or society generally
- Within a context
  - Primary data are uninterpreted
  - Data interpretation results in meaningful data
  - Meaningful data is information
- Primary data thus evolve to become contextually meaningful information
- Information about the natural and human worlds of interest
- Advance understanding for how observational data evolve to information
- A platform for discussion and advancement on this subject matter

Status Update since Montreal (P10)

- Developed and submitted Charter
- Obtained TAB review
- Obtained RDA endorsement
- Regular monthly meetings
- What started at P8 in Denver with a BoF is now an IG
- Clap, clap, clap ;>

Charter Overview

- Motivation
  - Frequent reference to the idea that information (knowledge) can be gained from data
  - By various people, infrastructures, projects, etc. (including RDA P11!)
  - Broad agreement this is true
  - Little agreement on how this occurs and what data and information (knowledge) are
- Specific concerns
  - Socio-technical support for the extraction of information from primary data
  - Systematic acquisition and curation of formal meaning of data
  - Construction and maintenance of information and knowledge-based systems
  - Further processing and use of information

Charter Overview: Objectives

- Identify, possibly develop, a reference conceptualization
  - Ground our understanding of the distinction of observational data and information
  - As well as the relevant activities and agents in between
- Engage stakeholders
  - Research communities, including individual researchers and ICT specialists
  - Research infrastructures, data infrastructures, data centers, e-Infrastructures
  - Other relevant RDA groups
  - Learn from a wide range of communities and practices
  - Devise solutions that are viable and practical across stakeholders
- Collect comparable use cases, solutions and challenges
  - Analyse use cases and develop solutions for unresolved challenges
  - Transfer solutions across stakeholders

Charter Overview: Outcomes

- Systematic acquisition of information by infrastructures
- Infrastructure to support data use as-a-service
- Information systems layered above current data systems
- Improved usability of data as information by both humans and machines

TAB Review (Positive)

- Very comprehensive charter and summary
- Well described demonstrating a sufficient expertise of the authors
- Topic well aligned with the RDA mission
- Worthwhile IG that is likely to add value to what is currently being done
- Outcomes are likely to lead to more meaningful data sharing and exchange

TAB Review (Improvements)

- Expansion of the membership, both geographically and in discipline expertise
- References to activities in other continents are missing
- Further external organizational outreach
- Involve GEO BON and aerosol scientists (for use cases)
- Number discrepancy between those who signed the charter and signed up

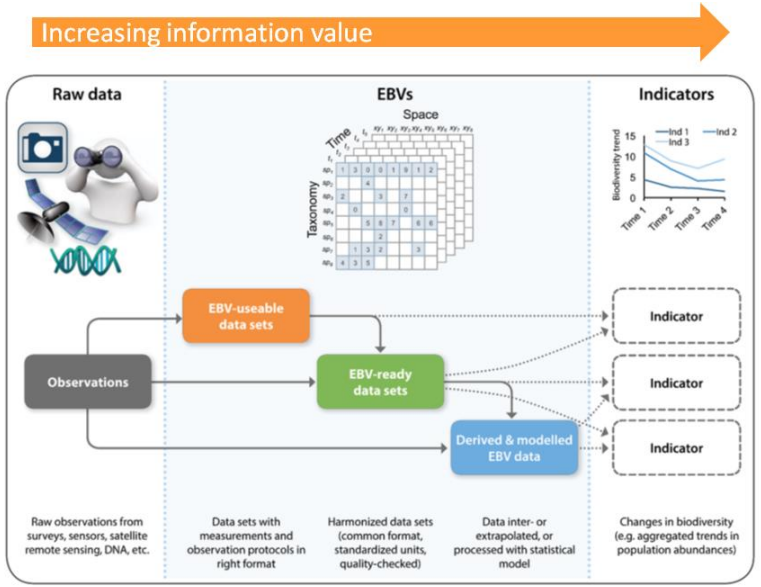
IDW session

“From Data to Knowledge: A Policy Perspective"

Biodiversity & Conservation Science: Summary

Essential Biodiversity Variables (EBVs) are conceptually positioned between raw data (i.e. primary data observations) and indicators (synthetic indices for reporting change)

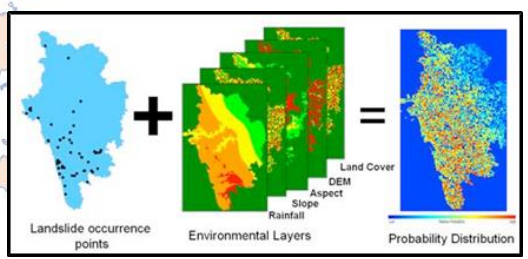
Information for a purpose: Understanding and reporting biodiversity change (science, policy, management)



Observational data: Structured primary biodiversity observations (EBV-useable data)

Information: EBV-ready data permit: i) analysis of, for example invasiveness; ii) other derived information products

Activity: Interpreting EBV-usable and EBV-ready data with expert knowledge and statistical models



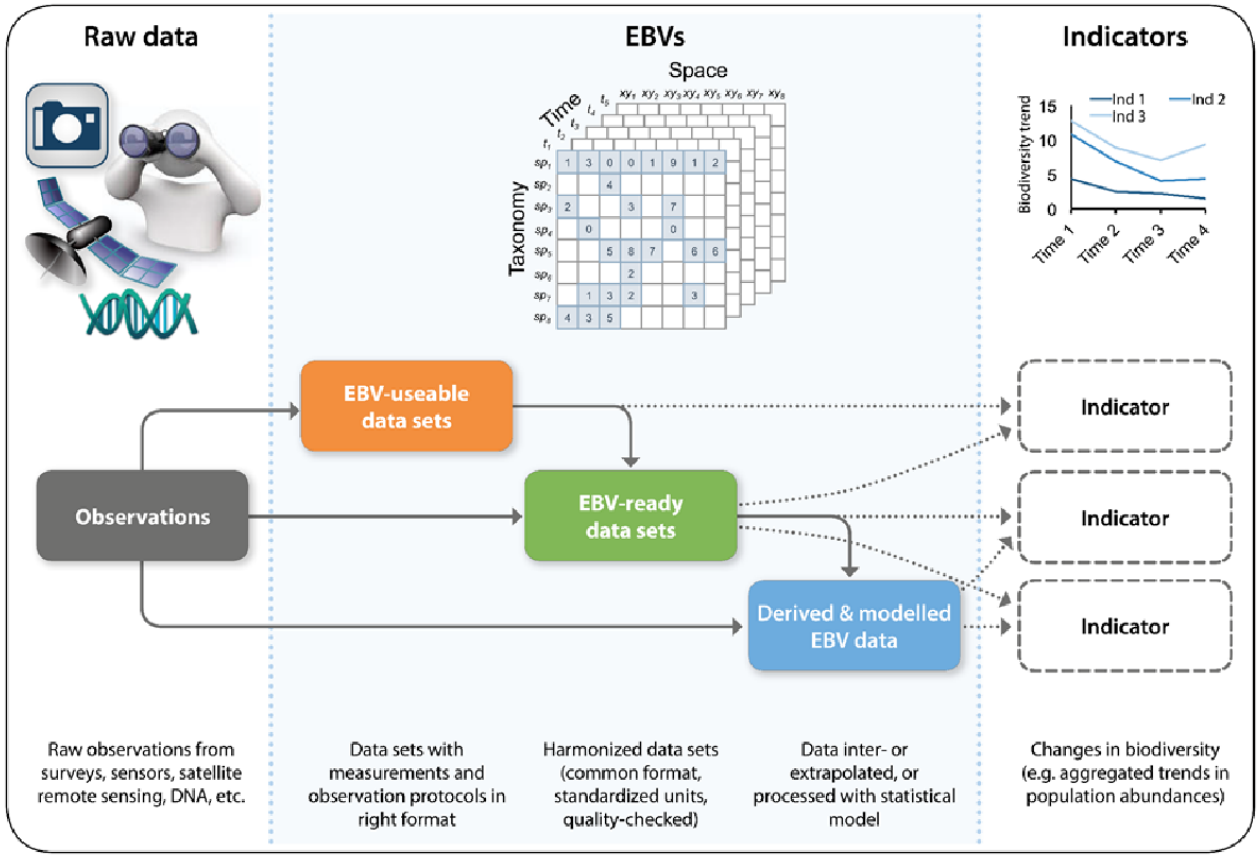
Essential Biodiversity Variables for species distribution and abundance

A Use Case in Biodiversity and Conservation Science

(use case document: https://goo.gl/U98Tj8
article: Kissling et al. 2018, doi: 10.1111/brv.12359)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No

What are EBV's

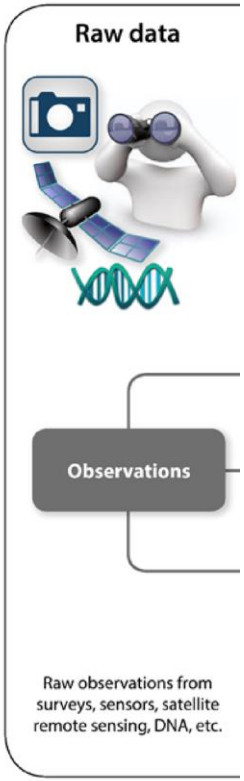


- Essential Biodiversity Variables (EBVs) are part of an information supply chain, conceptually positioned between raw data (i.e. primary data observations) and indicators (synthetic indices for reporting change)
- Information for a purpose: Understanding and reporting biodiversity change (science, policy,

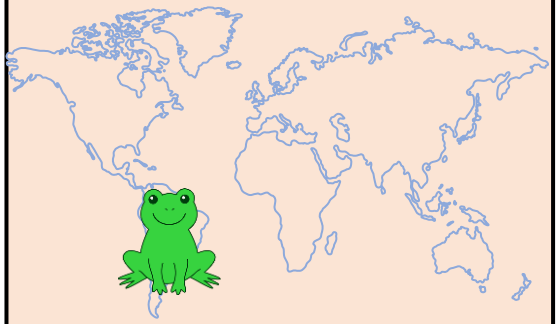
Observations / primary data

Measurements and observations in many formats

Surveys, sensors, satellites, DNA, etc.



Example:

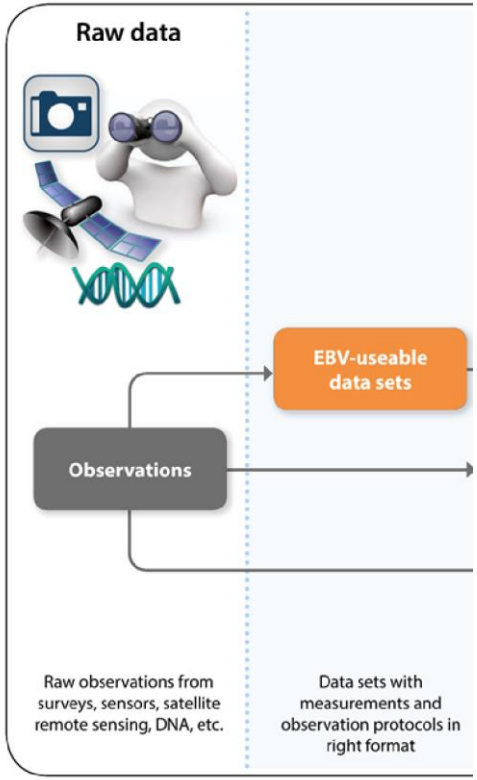


Raw observation data records presence of a species at a specific geographical location at a specific point in time

Clipart from http://www.clipartpanda.com/,

1) Observations / primary data to EBV usable data

Measurements with comparable units, similar observation protocols



When raw data is structured, well-formed, based on comparable measurement units using similar observation protocols, it is usable for producing EBV data products

Activities

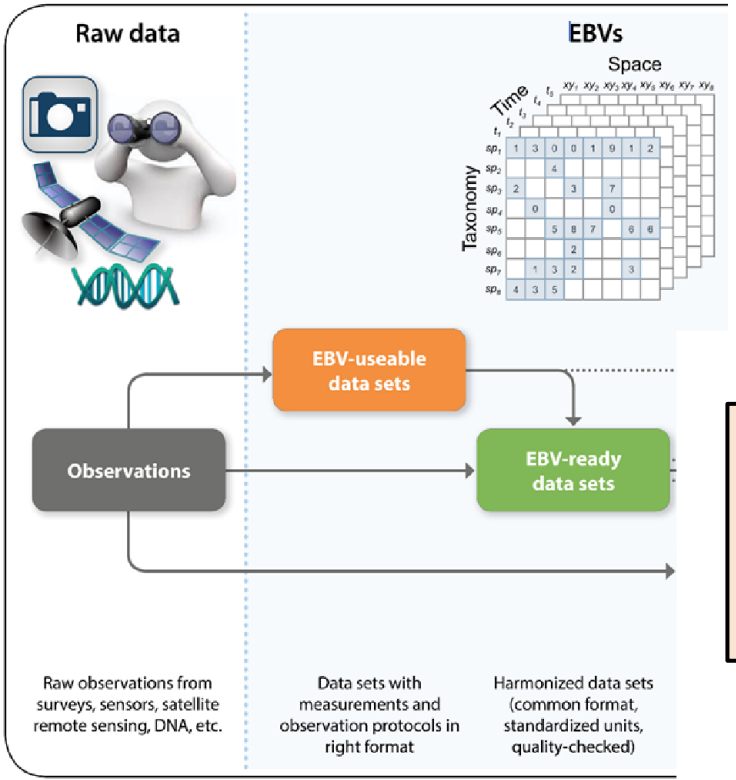
- Discovery and retrieval from repositories
- Filtering by key dimensions of taxonomy, time and space
- Structuring and formatting

Involves applying expert knowledge and judgement

2) EBV usable data to EBV ready data

Harmonised datasets, common format, standardized units, quality-checked

Structuring, well-forming, packaging, adding 3rd-party detail



EBV ready data are usable information objects. They possess sufficient context and meaning

Activities

Assessing scientific compatibility and technical interoperability of data

Assessing legal interoperability of data (open access, licensing restrictions)

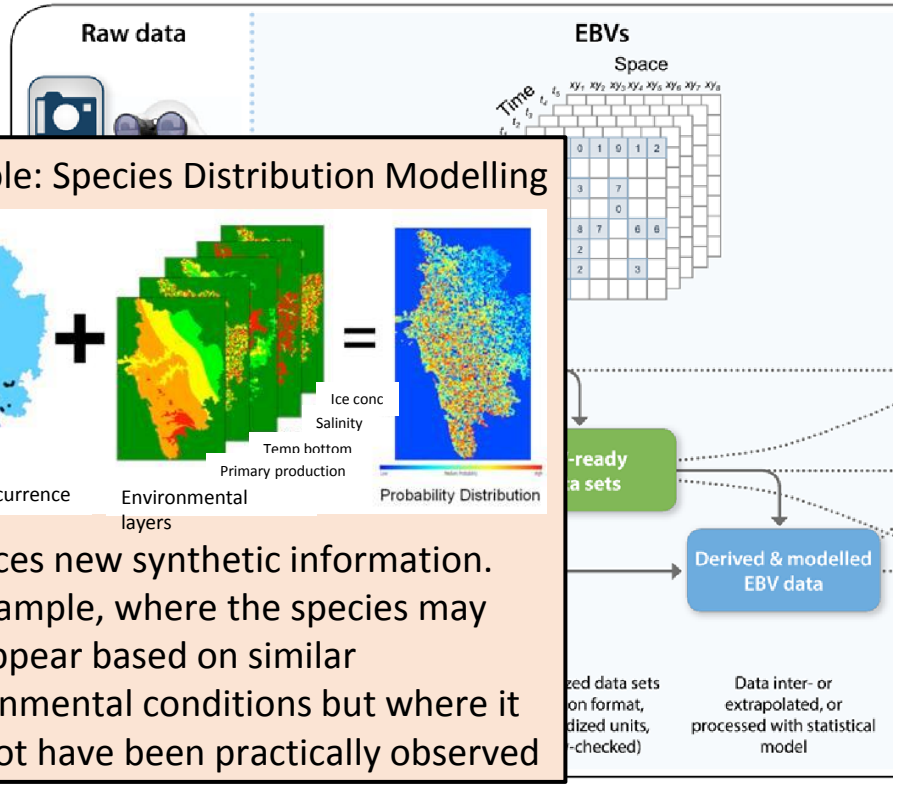
Applying quality control procedures and adding assertions e.g., on accuracy of geographical information; removing duplicates

Combines automation with expert human judgement

3) EBV ready data to derived & modelled EBV data

Derived from processing data with statistical models

Interpretational processing, modelling, etc.



Example: Species Distribution Modelling

Produces new synthetic information. For example, where the species may also appear based on similar environmental conditions but where it may not have been practically observed

Activities

Increasingly complex processing with higher level of human expert input also often needed

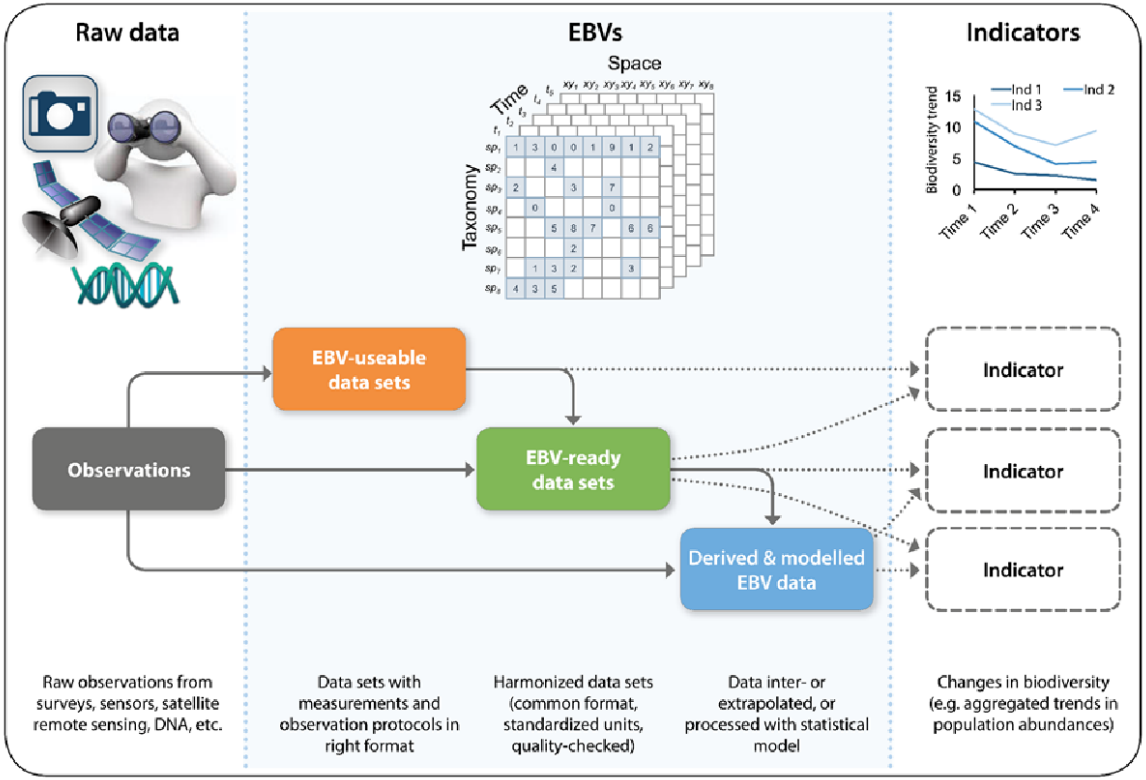
Recording processing steps (i.e., provenance), both human and machine readable

Derived & modelled EBV ready data can be used for gap-filling. They are also usable information objects

4) EBV data to indicators

e.g., quantifying spatiotemporal changes in distributions / abundances

Synthesised from multiple sources by processing and interpretation



Activities

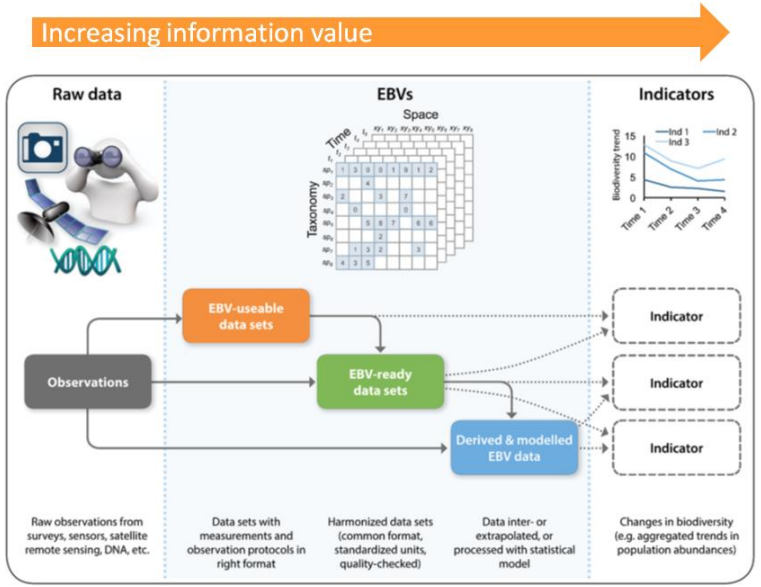
Synthesising indicators relevant to e.g., Aichi 2020 Biodiversity Targets, Sustainable Development Goals 2030, etc.

Quantifying uncertainty arising from combining data acquired by different methods

Biodiversity & Conservation Science: Summary

Essential Biodiversity Variables (EBVs) are conceptually positioned between raw data (i.e. primary data observations) and indicators (synthetic indices for reporting change)

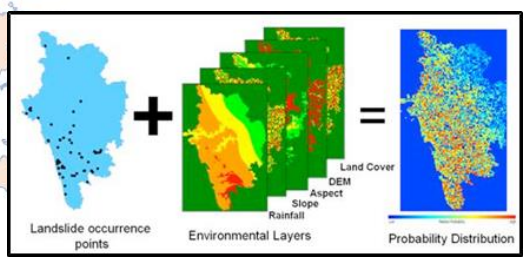
Information for a purpose: Understanding and reporting biodiversity change (science, policy, management)



Observational data: Structured primary biodiversity observations (EBV-useable data)

Information: EBV-ready data permit: i) analysis of, for example invasiveness; ii) other derived information products

Activity: Interpreting EBV-usable and EBV-ready data with expert knowledge and statistical models



GLOBIS-B

Acknowledge global cooperation

Project partners:

- University of Amsterdam, NL
- Cardiff University, UK
- Gnubila, FR
- National Research Council, IT
- University of Alcala, ES
- Martin-Luther University Halle-Wittenberg, DE

ATLAS OF LIVING AUSTRALIA
sharing biodiversity knowledge

中国科学院生物多样性委员会
Biodiversity Committee,Chinese Academy of Sciences

DataONE
Data Observation Network for Earth

cria

elixir

GBIF

WDCM

GEO BON

Kunming Institute of Botany, Chinese Academy of Sciences

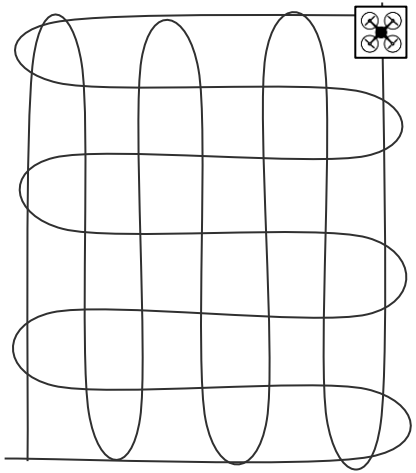
LifeWatch

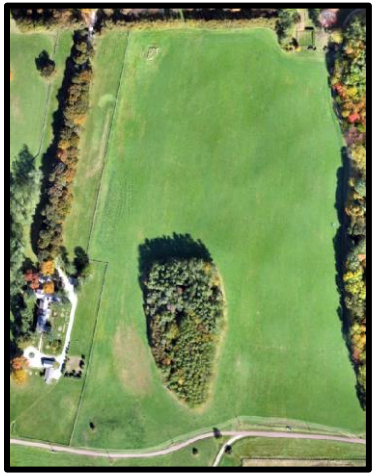
neon
National Ecological Observatory Network, Inc.

SANBI
Biodiversity for Life
South African National Biodiversity Institute

Example: Scientific Unmanned Aircraft Systems

- Observational data: Multispectral Imagery
- Information: Manure Nutrient Management and Biomass Estimations
- Activity: Evaluation of agricultural soil climate change mitigation potential





Equation 7. Emission Equation for Direct N_2O Emissions from Agricultural Soils

Emissions ($MMTCO_2E$) =
Total N × fraction unvolatilized (0.9 synthetic or 0.8 organic)
× 0.01 (kg N_2O-N/ kg N) × 44/ 28 (Ratio of N_2O to N_2O-N) × 298 (GWP)
÷ 1,000,000,000 (kg/ $MMTCO_2E$)

Agriculture	1.0	0.96	0.92	0.84	0.84	0.84
Enteric Fermentation	0.59	0.56	0.53	0.50	0.53	0.53
Manure Management	0.12	0.14	0.15	0.17	0.16	0.16
Agricultural Soils	0.29	0.26	0.24	0.17	0.15	0.15
TOTAL GROSS EMISSIONS	8.11	8.86	9.34	8. 23	8.11	8.27
Change relative to 1990	-	+9%	+15%	+1.5%	0%	+2%

Precision Agriculture

- Observational data: Weather data including temperature and humidity

Precision Agriculture

- Observational data: Weather data including temperature and humidity
- Information: Descriptions for situations of (acute) outbreaks of pests in crops

Precision Agriculture

- Observational data: Weather data including temperature and humidity
- Information: Descriptions for situations of (acute) outbreaks of pests in crops
- Activity: Forecast disease pressure using a physically based model

Intelligent Transportation Systems

- Observational data: Road pavement vibration

Intelligent Transportation Systems

- Observational data: Road pavement vibration
- Information: Descriptions of vehicles, their type, speed and driving direction

Intelligent Transportation Systems

- Observational data: Road pavement vibration
- Information: Descriptions of vehicles, their type, speed and driving direction
- Activity: Machine learning classification of vibration patterns

Work Plan

- OD2I IG kick-off session at Plenary 11 in Berlin
- Liaise with related RDA groups, and groups outside RDA (e.g. GEO/GEOSS)
- Develop the OD2I IG’s reference conceptualization
- White paper on developed reference conceptualization
- Collect new use cases and align them with the reference conceptualization
- Analyse the use cases for commonalities and differences
- Identify and report common challenges
- Collect feedback from teams implementing use cases

Discussion

- What do the presented use cases have in common
- How to expand the membership
- Collaborations with other groups at RDA (e.g. VRE IG)
- New use cases proposed by audience
- Relevant activities in other continents
- Conceptual frameworks to consider