

From observational data to information IG

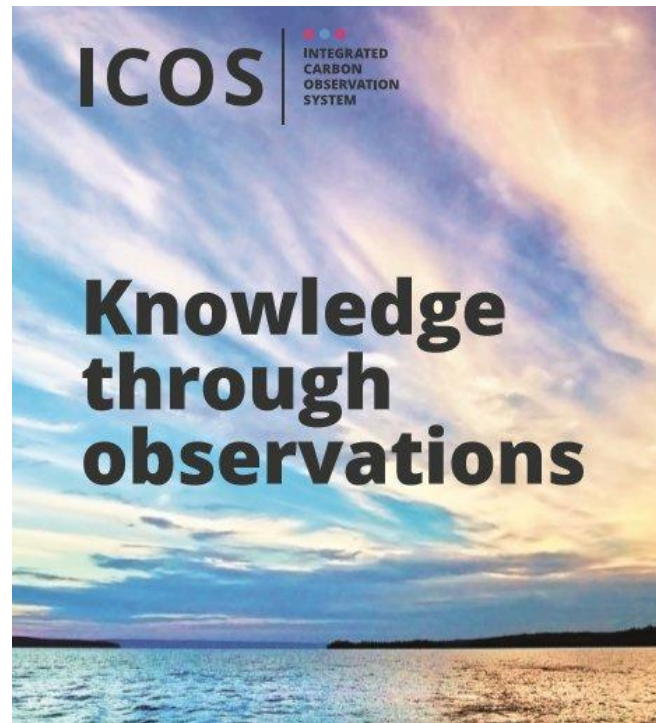
Markus Stocker, Jay Pearlman, Stefano Nativi, Ari Asmi
Jacco Konijn, Alex Hardisty *and the IG*

bit.ly/2xadQsf

Collaborative session notes

About

- Relationship between data and information
- Observational data
- Semantic information about the environment
- Environmental research infrastructures



Why

- Common ideas
 - Mining information from data
 - Transfer of information into knowledge
 - Research data for better decisions
 - Actionable information/knowledge
- But what does this mean
- Information about what
- What are relevant processes
- How does infrastructure support this
- Is information actionable for infrastructures, or just human experts
- ...

History

- It all started at P8 in Denver
- BoF organized by Ari Asmi, Stefano Nativi, Jay Pearlman, Peter Wittenburg
- Lunch meet-up at AGU 2016
- Second BoF at P9 in Barcelona

Outlook

- Critical milestone is drafting the Charter
- Planned for P11 in Berlin next Spring
- Attain RDA endorsement

rd-alliance.org/groups/observational-data-information

obs-data-info@rda-groups.org

Update on activities since P9

- Settled on IG, rather than WG
- Decided IG name “From observational data to information”
- Regular monthly calls, first Monday of the month, 4-5 pm (Berlin)
- Work on comparable use cases, based on template
- Currently one on biodiversity indicators and one in aerosol science
- Setup RDA web pages
- Setup RDA mailing list
- Setup Google Drive folder for document management/collaboration



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. 2017, doi: [10.1111/brv.12359](https://doi.org/10.1111/brv.12359))

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654003.





Funding

EU-funded project, Horizon 2020

Call:

International cooperation for research infrastructures

Type of action:

Coordination and support action

Duration:

3 years (June 2015 to May 2018)

Funding:

1 M euro

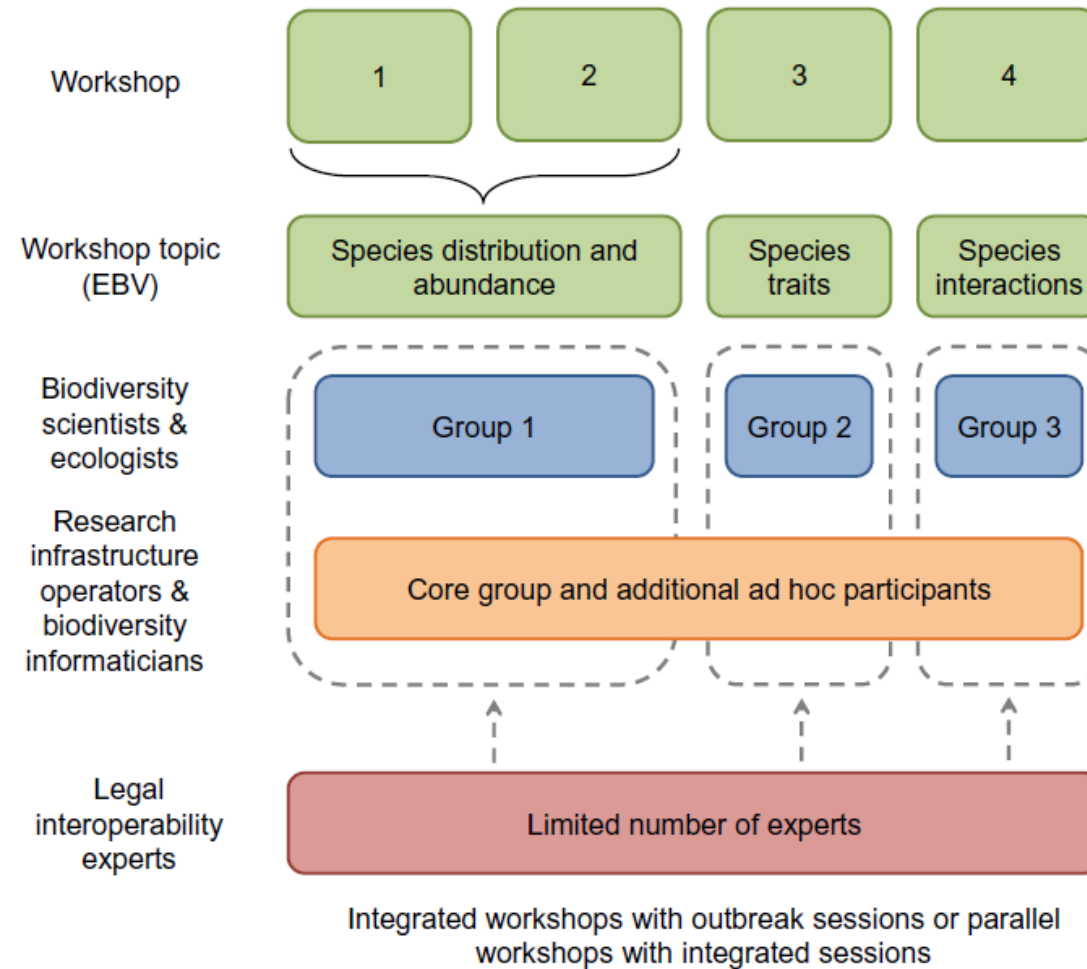


Global Cooperation





Workshops





Lead partners

- Dr. W. Daniel Kissling, Associate Professor for Quantitative Biodiversity Science at the Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam.
- Alex Hardisty, Director of Informatics Projects in the School of Computer Science and Informatics, Cardiff University.
- Prof. Enrique Alonso, Legal Counselor, Consejo de Estado, Spain.
- Jacco Konijn, Head of Project Management, University of Amsterdam.

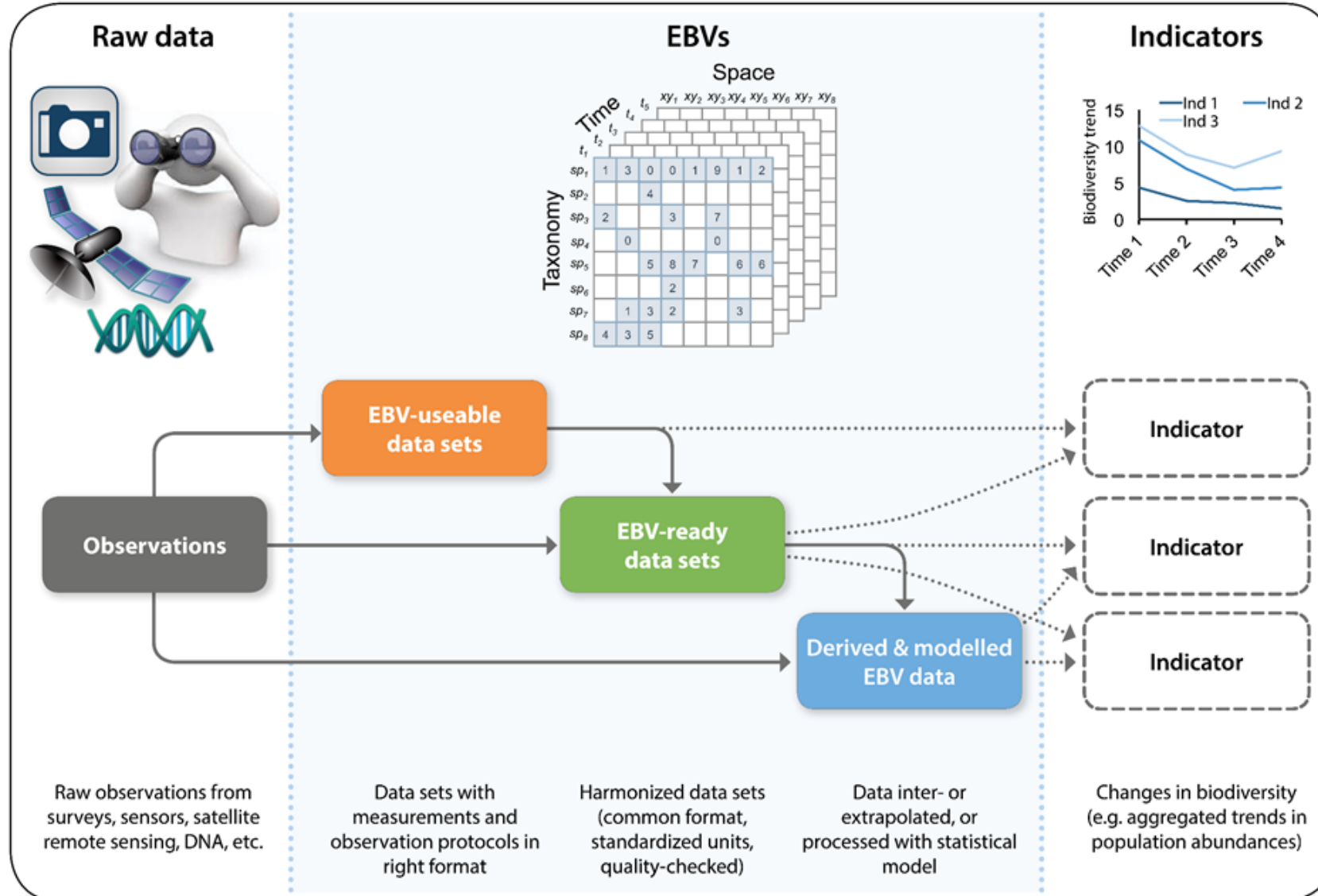


What are EBV's

- 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, management)



Increasing information value

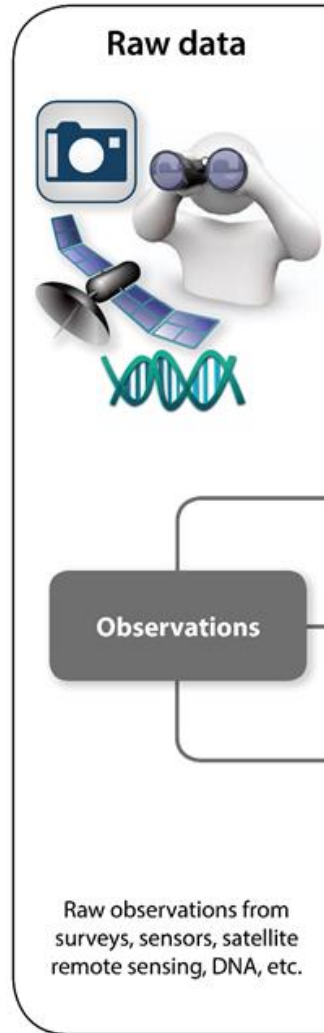




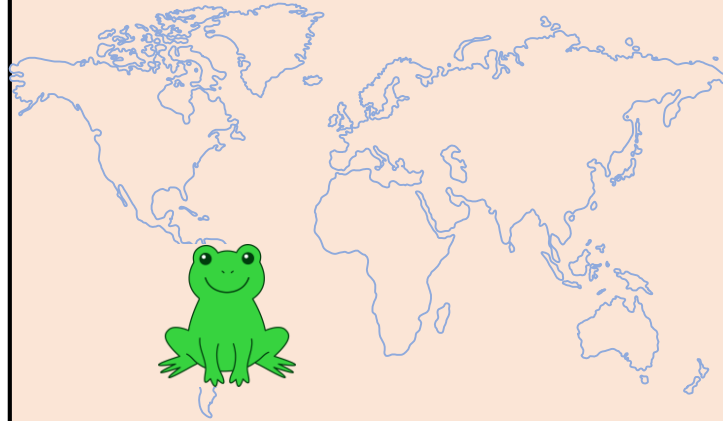
Observations / primary data

Measurements and observations in a variety of formats

Surveys, sensors, satellites, DNA, etc.



Example:



Raw observation data from multiple sources records the presence of a species at a specific geographical location at a specific point in time

Issues / requirements

Sufficient and adequate metadata

Observations / primary data to EBV usable data

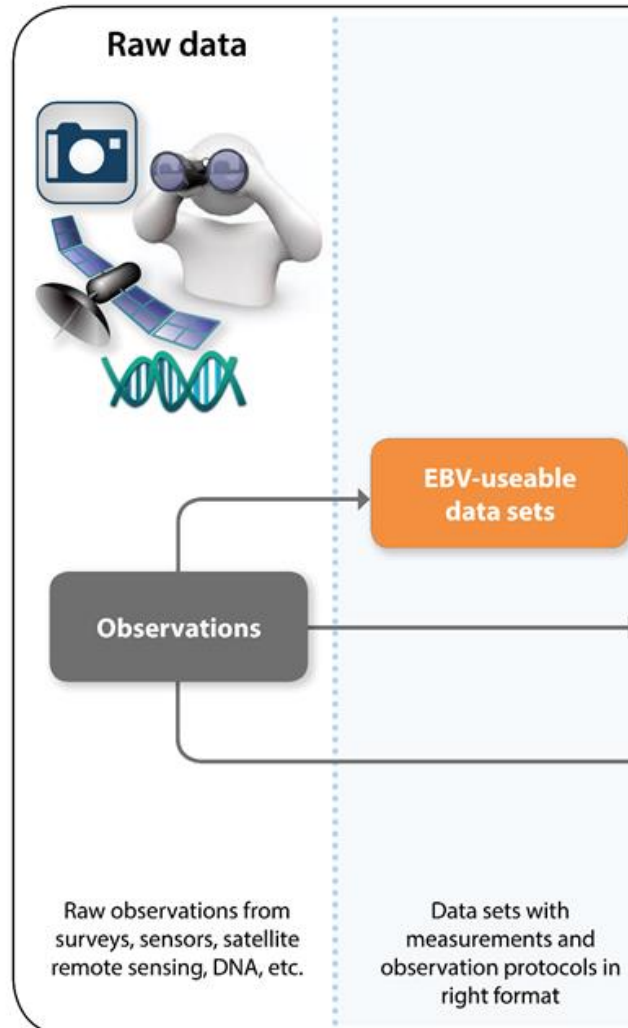
Measurements with comparable units, similar observation protocols

Issues / requirements

Discovery and retrieval of available relevant observations from data repositories

Filtering by key dimensions of taxonomy (species), time and space

Requiring expert knowledge and judgement



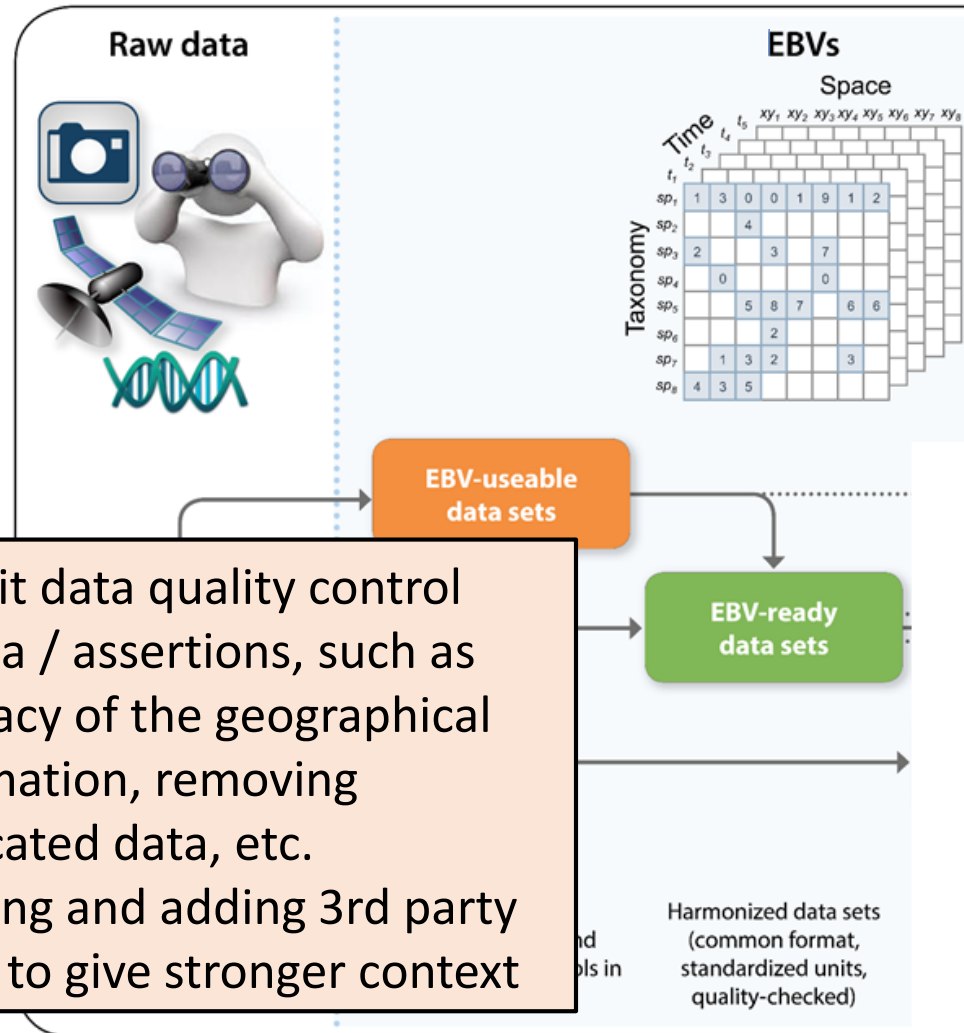
When raw data is structured, well-formed, based on comparable measurement units using similar observation protocols, it is usable for producing EBV data products

EBV usable data to EBV ready data

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

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

Issues / requirements



Explicit data quality control criteria / assertions, such as accuracy of the geographical information, removing duplicated data, etc.
Merging and adding 3rd party detail to give stronger context

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

Agreement on processing steps

Scientific compatibility and technical interoperability of data

Legal interoperability of data (i.e., open access, removal of licensing restrictions)

Sufficient and harmonised metadata

Harmonisation of QC approach

Combining automation and expert human judgement

Structural standards missing

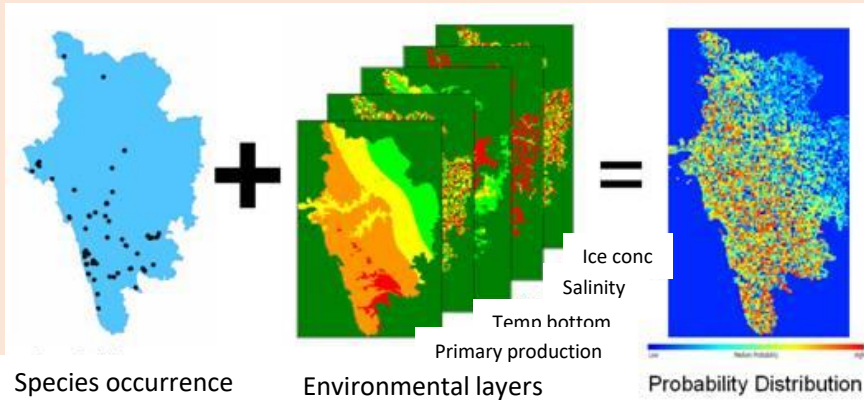
EBV ready data to derived & modelled EBV data

Derived from processing data with statistical models

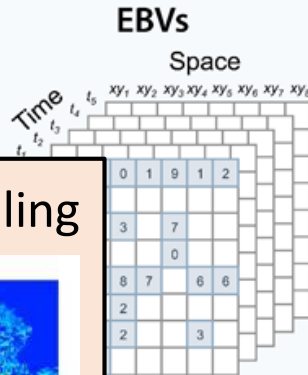
Interpretational processing, modelling, etc.

Issues / requirements

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



ready
a sets

Derived & modelled
EBV data

ized data sets
on format,
dized units,
-checked)

Data inter- or
extrapolated, or
processed with statistical
model

Increased complexity

Automation more beneficial but higher level of human expert input also often needed

Transparent record of processing steps (i.e., provenance), both human and machine readable

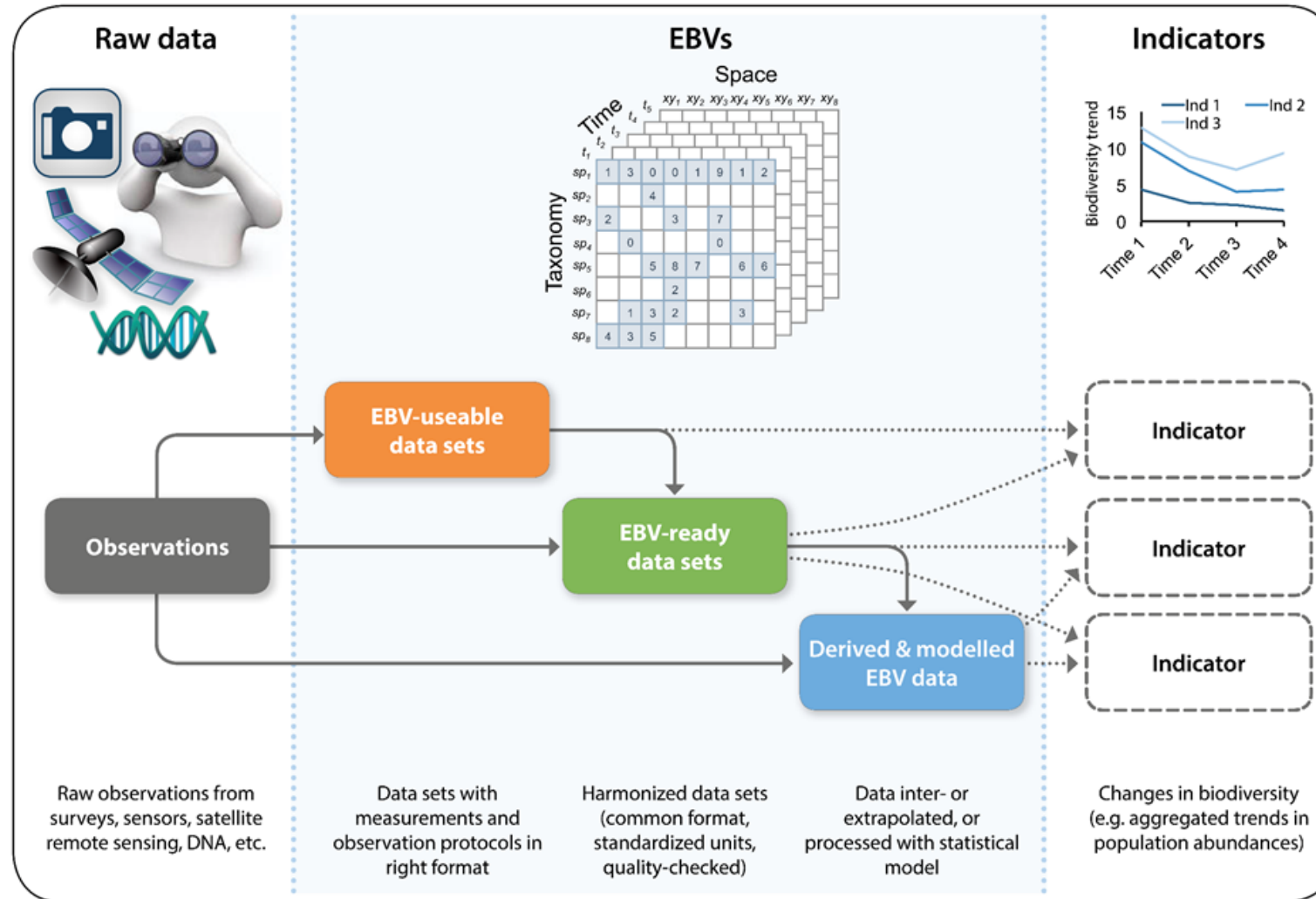
Derived & modelled EBV ready data can be used for gap-filling. They are also usable information objects

EBV data to indicators

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

Synthesised from multiple sources by processing and interpretation

Issues / requirements

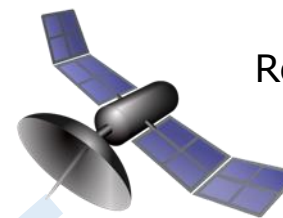


Indicators must be relevant e.g., to Aichi 2020 Biodiversity Targets, Sustainable Development Goals 2030, etc.

Basis of an indicator must be clear so that repeated assessments over time are possible

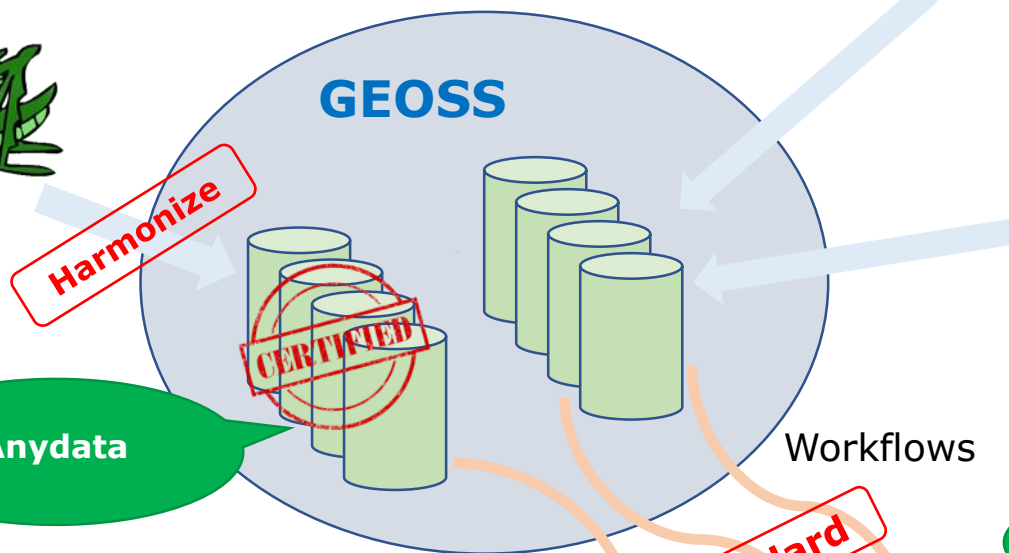
Quantifying uncertainty arising from combining data acquired by different methods

Methods evolving over time



Remote Sensing

Modelled data/algorithms



Drivers and Pressures



Anyone

Standard

Anything

Metagenomics/
DNA data



Standard

Anytime

**EBV's and
indicators for
GEO BON**

Anywhere

Other use cases

Aerosol science

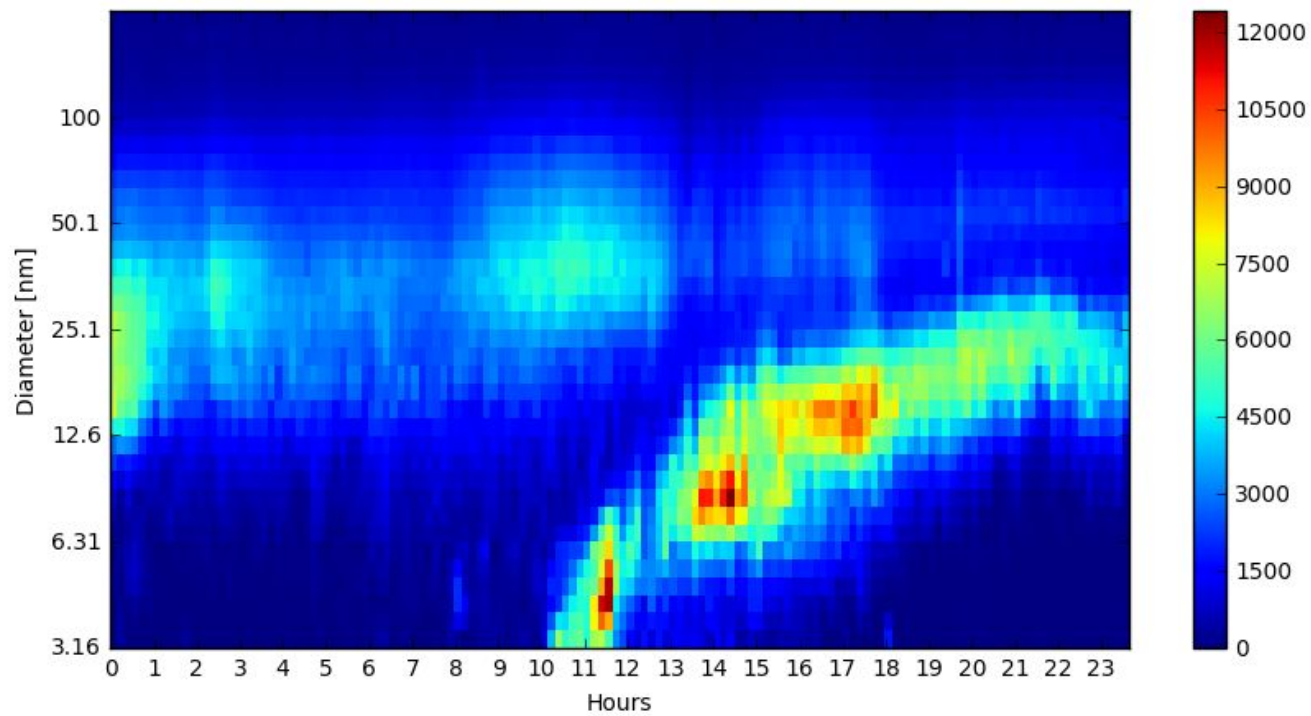
Intelligent transportation systems

Disease outbreaks in agriculture

Pattern

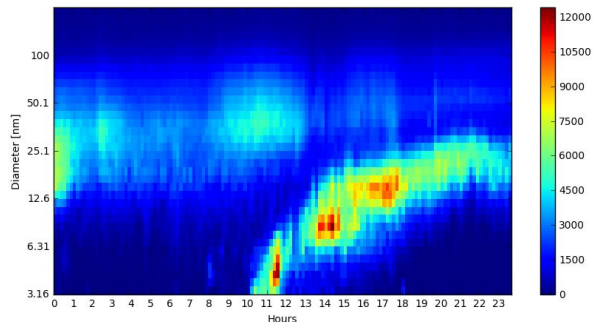
- Primary observational (sensor) data
- Data interpretation
- Derived information about observed environment
- **Information is formal (machine readable)**

Aerosol science



```
from smear.datafetcher import fetchdata
from smear.dataplotter import plotdata

# Fetch and plot concentration data for the given time and location
# from SmartSMEAR, https://avaa.tdata.fi/web/smart
plotdata(fetchdata('2013-04-04', 'Hyytiälä'))
```



```
from smear.datafetcher import fetchdata
from factory import assess

# Automated assessment for whether or not an event occurred
assess(fetchdata('2013-04-04', 'Hyytiälä'))
```

```
['Event']
```

```
from factory import record, event

# Record information about the new particle formation event
record(event('2013-04-04', 'Hyytiälä', '11:00', '19:00', 'Class Ia'))
```

<http://pangaea.de/eb1ad69f11aecd2449f6d5741c3b8ac3> a lode:Event ;
 smear:hasClassification smear:ClassIa ;
 lode:atPlace <http://sws.geonames.org/656888/> ;
 lode:atTime <http://pangaea.de/92be5465a05cc56156422d6cdb4603e1> ;
 lode:inSpace <http://pangaea.de/7f885190eb43154e01c97f814b287a4b> .

<http://pangaea.de/3ca2edb9bb13c998a0ea94082aa592dd> a time:Instant ;
 time:inXSDDateTime "2013-04-04T11:00:00+03:00"^^xsd:dateTime .

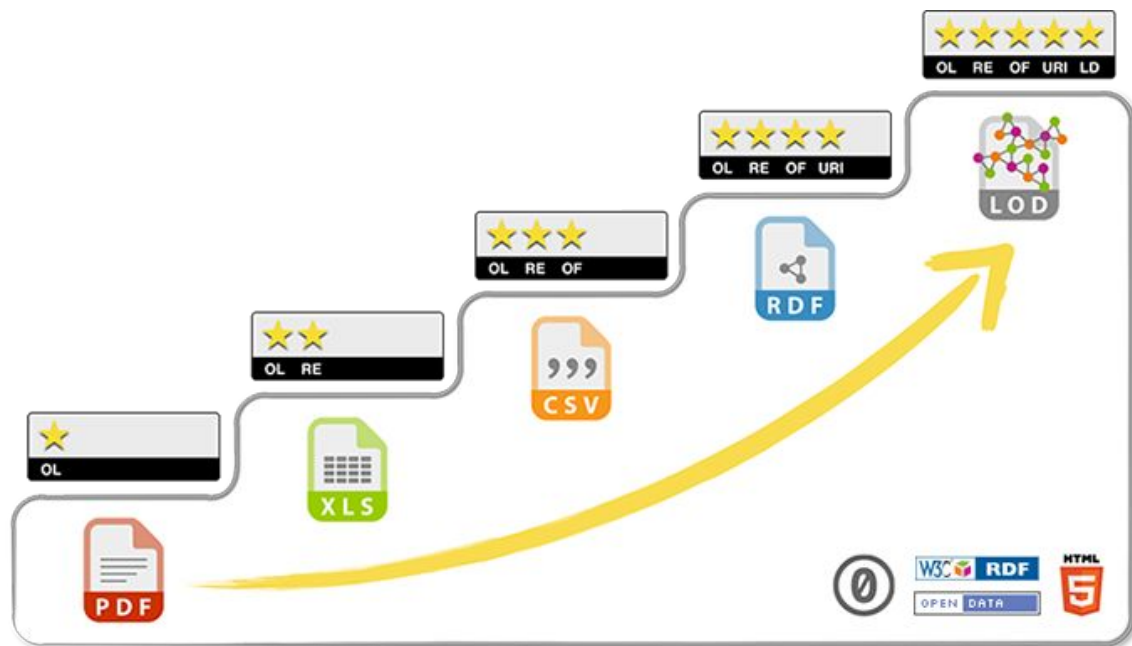
<http://pangaea.de/7f885190eb43154e01c97f814b287a4b> a sf:Point,
 wgs84:SpatialThing ;
 geosparql:asWKT "POINT (24.29077 61.84562)"^^geosparql:wktLiteral .

<http://pangaea.de/92be5465a05cc56156422d6cdb4603e1> a time:Interval ;
 time:hasBeginning <http://pangaea.de/3ca2edb9bb13c998a0ea94082aa592dd> ;
 time:hasEnd <http://pangaea.de/cc87885836375b64bf023a6bcd5ad814> .

<http://pangaea.de/cc87885836375b64bf023a6bcd5ad814> a time:Instant ;
 time:inXSDDateTime "2013-04-04T19:00:00+03:00"^^xsd:dateTime .

<http://sws.geonames.org/656888/> a gn:Feature,
 DUL:Place ;
 gn:countryCode "FI"^^xsd:string ;
 gn:locationMap <http://www.geonames.org/656888/hyytiaelae.html> ;
 gn:name "Hyytiälä"^^xsd:string ;
 wgs84:lat "61.84562" ;
 wgs84:long "24.29077" .

734544, 11:00, 19:00, Classla, Hyytiälä



<http://5stardata.info/en/>

Intelligent transportation systems

- Detection of vehicles using road-pavement vibration
- Several vibration sensors (accelerometers) installed in road pavement
- Observational data
 - Road pavement vibration (acceleration)
- Data interpretation
 - Classification of vibration patterns
- Derived information
 - About detected vehicles
 - Type, speed, driving direction

Disease outbreaks in agriculture

- Describe situations of disease outbreak in crops
- Diseases are fungal pathogens
- Observational data
 - Weather data such as humidity, temperature, wind speed
- Data interpretation
 - Computation of cumulative disease pressure
 - Using a disease pressure model
 - Parameterized with crop and tillage type
 - Executed daily on weather data
- Derived information
 - About situations of disease outbreak
 - Severity, duration, type of pathogen and crop, location

Update on Charter

Introduction

A brief articulation of what issues the IG will address, how this IG is aligned with the RDA mission, and how this IG would be a value-added contribution to the RDA community

- Clearly establish the rationale for why we need this group
- Review current understanding of the differences between data, information, knowledge
- Focus on the process, value chain, more than on the entities
- ...

Update on Charter

User scenario(s) or use case(s) the IG wishes to address

What triggered the desire for this IG in the first place

- We have something to show here
- Though IG may wish to address different use cases
- Contribute your use case
- ...

Update on Charter

Objectives

A specific set of focus areas for discussion, including use cases that pointed to the need for the IG in the first place. Articulate how this group is different from other current activities inside or outside of RDA.

- Better grasp for what “data to information” means
- Focus on data use phase of research data lifecycle
- What happens on the interface between infrastructures and research communities
- Latter part relies on some kind of landscaping
- ...

Update on Charter

Participation

Address which communities will be involved, what skills or knowledge should they have, and how will you engage these communities. Also address how this group proposes to coordinate its activity with relevant related groups.

- Research infrastructures
- Research communities
- ICT
- ...

Update on Charter

Outcomes

Discuss what the IG intends to accomplish. Include examples of WG topics or supporting IG-level outputs that might lead to WGs later on.

- White papers
- Demonstrators
- Mapping the landscape as a deliverable
- ...

Update on Charter

Mechanism

Describe how often your group will meet and how will you maintain momentum between Plenaries

- This is fairly established
- Monthly conference calls
- Calls as status updates, assigning actions

Update on Charter

Timeline

Describe draft milestones and goals for the first 12 months

- Should be straightforward once the objectives are clear
- ...

Update on Charter

Potential Group Members

Include proposed chairs/initial leadership and all members who have expressed interest

- There are proposed chairs but not set in stone

- What do we want to get out of this IG
 - Motivation, goals, intentions, outputs
- How to clarify the difference between *observational* data and information
- Develop the Charter
 - Key task for the next six months
- Use case contributions
 - Bottom up activity
 - Agree on template!
- Relationships with other RDA groups: Mapping of the Landscape
 - Include this as a deliverable of the IG
 - Feedback into Atlas of Knowledge
- IG to spawn WGs that tackle concrete challenges
- Are you joining this group, what is your motivation?
- Do you see overlaps in your work with what was presented?
- Public meeting notes (archived at RDA pages)? Yes/No