Ten outcomes to improve informatics interoperability in cyber/e-Infrastructures for biodiversity and ecological sciences

(through the use case of Essential Biodiversity Variables)

GLOBIS-B

GLOBal Infrastructures for
Supporting Biodiversity research

Alex Hardisty
School of Computer Science & Informatics
email: hardistyar@cardiff.ac.uk
/alexhardisty

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Project partners

Project partners

University of Amsterdam

Cardiff University

Prifysgol Caerdydd

Universidad de Alcalá

Consiglio Nazionale delle Ricerche

cnrs

comunità ambiente

gnúbila

GEO BON

Other collaborators

Monash University

CSIRO

Research infrastructures

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

SANBI Biodiversity for Life
South African National Biodiversity Institute

DataONE
Data Observation Network for Earth

GBIF
Global Biodiversity Information Facility

LifeWatch

elixir
European infrastructure for biological information

WDCM
World Data Centre of Microorganisms

DiSSCo
Distributed System of Scientific Collections

Atlas of Living Australia
sharing biodiversity knowledge

Kunming Institute of Botany, Chinese Academy of Sciences
Germplasm Bank of Wild Species

LTER

cria
Brazilian Reference Center on Environmental Information

neon
National Ecological Observatory Network, Inc.

Coordination & Support Action

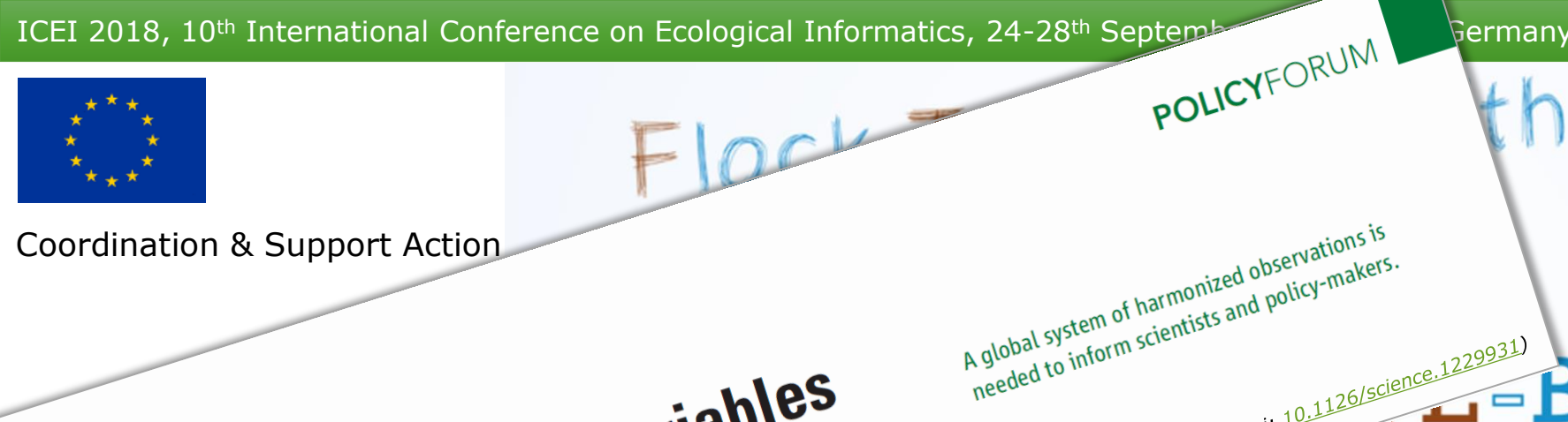
How to improve informatics interoperability between research infrastructures for biodiversity and ecological sciences?

Flock Together with

CReATIVE-B

A roadmap of global research data infrastructures supporting biodiversity and ecosystem science

Coordination & Support Action



Use case: Building Essential Biodiversity Variable (EBV) data products with global coverage and comparability.

Interested in infrastructures services that enable EBVs

What are EBVs?

'a minimum set of essential measurements to capture major dimensions of biodiversity change'

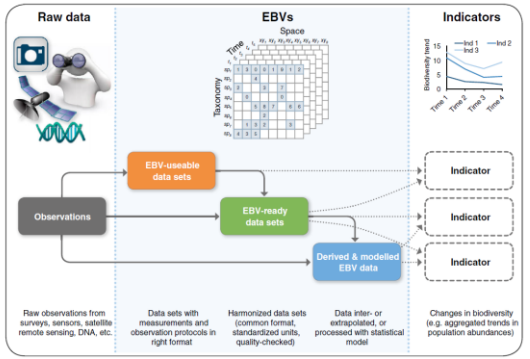
Since 2013, operational definitions, with work coordinated by GEO BON

22 variables proposed in 6 classes:

- Genetic composition
- Species populations
- Species traits
- Community composition
- Ecosystem function
- Ecosystem structure

EBVs are:

- Biological state variables measurable at specific points in time, space and biological organization.
- Harmonized measurements conceptually located between primary data observations ('raw data') and synthetic/derived indices ('indicators').



Brummitt 2017 (doi: 10.1016/j.biocon.2016.09.006); Navarro 2017 (doi: 10.1016/j.cosust.2018.02.005); Schmeller 2017 (doi: 10.1007/s10531-017-1386-9); Kissling 2018a (doi: 10.1111/brv.12359); Kissling 2018b (doi: 10.1038/s41559-018-0667-3); … … and several more

More info: https://geobon.org/ebvs/what-are-ebvs/

From informatics perspective, the main question:

How can we measure and calculate Essential Biodiversity Variables (EBVs) and prepare data products for them on a global scale?

How can it be done …

Anytime

Anywhere

… for any geographic area?

… at any relevant temporal scale?

Anything

… for any species, assemblage, ecosystem, or biome of interest?

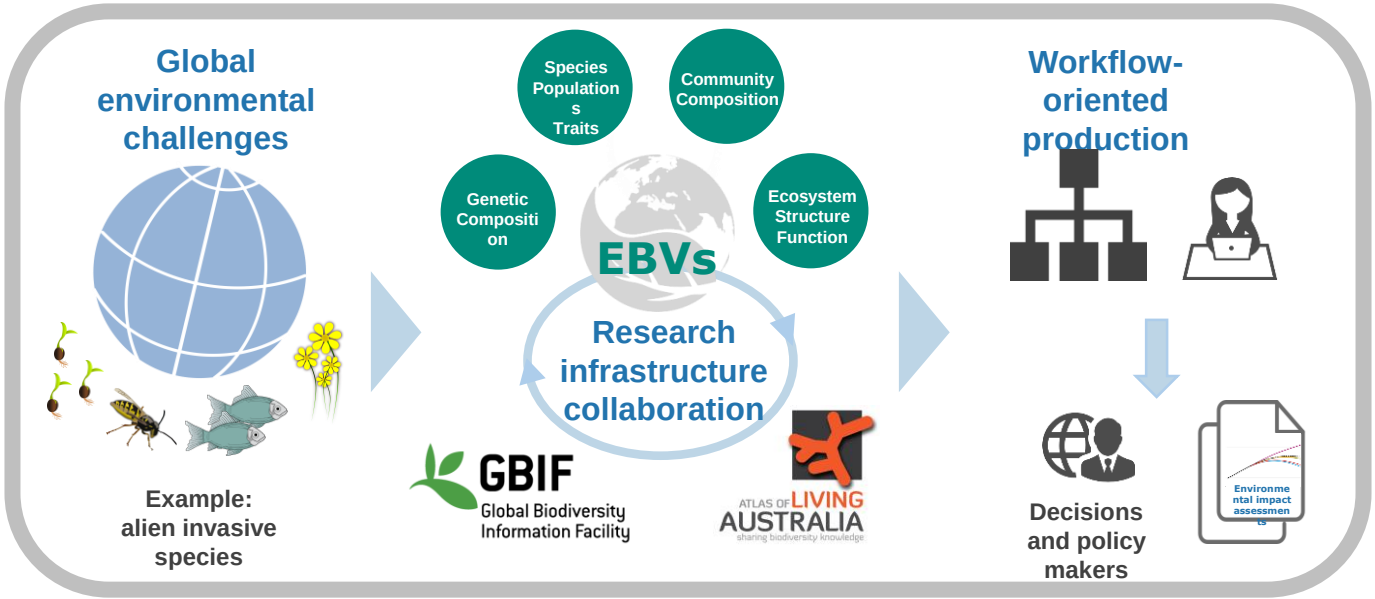
… with data that may be held by any (or across multiple) data repositories?

Anydata

… by any person anywhere in the world?

Anyone

Research infrastructure challenges in preparing Essential Biodiversity Variables data products for alien invasive species



Information on invasive alien species (bottom left) from (e.g.,) Global Invasive Species Database (GISD) and Global Register of Introduced and Invasive Species (GRIIS) combines with information from general research infrastructures such as GBIF and Atlas of Living Australia (centre) in a workflow to produce EBV data products (top right) that can be used by decision/policy-makers to tackle global environmental challenges.

Consistent quality checking across data from different sources

Common structure across EBV data and for packaging of products

Specific minimum information to make EBV data product usable

A workflow for the process to make it practical, deterministic and repeatable

A common, independent means of representing the workflow

Importance of recording what was done

'Bari Manifesto'

Ten guiding principles to enhance the ability of data and infrastructure providers to improve their support to production of global EBV data products

The Promenade, Bari, Italy

Principles 1 – 3

P1. Data Management Plan

Projects developing EBV data products should have a data management plan.

P2. Common standards for presenting data products

Common dimensions for all products (i.e., time, space, name/taxonomy, etc.).

Common file formats for processing by wide range of software.

P3. With human and machine-readable metadata

Both for the products themselves and for source data.

Based on community standards and sufficient for data discovery, access, fitness-for-purpose evaluation, interpretation and use, and citation.

Principles 4 – 6

P4. Quality-assured and fit-for-use

EBV data products and their component sub-parts should undergo quality testing and include data quality information sufficient to identify fitness-for-use of the data for specific purposes.

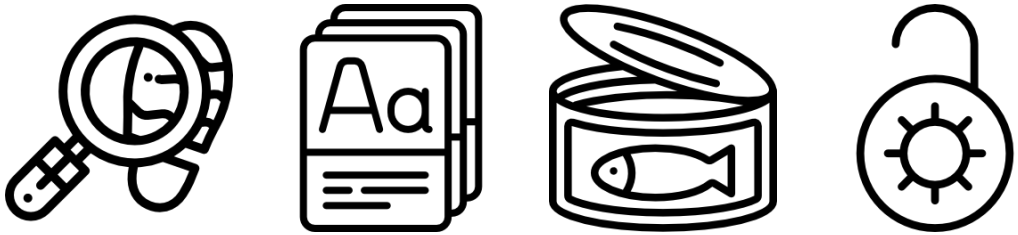
P5. Accessible via standardized Application Programming Interfaces

EBV data products should expose their capabilities and be accessible through common, standardized Application Programming Interfaces (APIs).

P6. Prepared using standard workflows

Workflows for preparing and publishing EBV data products must be fully documented and published. Must be repeatable. Ideally, represented in a non-proprietary manner, thus allowing execution/replication elsewhere.

Principles 7 – 10



P7. Traceable

Possible to trace process from product back to raw data and to reproduce the process. Provenance information readable by humans and by machines.

P8. Described using standard vocabularies

EBV data products should be described by standard, open, and machine-readable vocabulary terms and conceptual relations (ontologies).

P9. Preserved with a persistent identifier

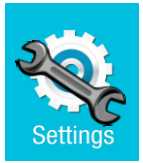
EBV data products and underlying raw data should be preserved with a persistent identifier in a community supported, open and trusted data repository.

P10. Open and 'FAIR'

Legally open, and 'Findable, Accessible, Interoperable and Reusable' (FAIR). For all assets – raw data, EBV data product, workflows, etc.

Finally, research infrastructures express their interest to cooperate further around ten 'Bari' principles to achieve success for routine generation of EBV data products on the global level.

'*a minimum set of essential measurements to capture major dimensions of biodiversity change*'



Manifesto is a step towards removing variations in the process – essential!

Bari Manifesto principles are generally and widely applicable to just about any sphere of ecological informatics. Adopting even one principle in your own work improves matters for everyone.

GLOBIS-B

GLOBal Infrastructures for Supporting Biodiversity research

The work has been performed by the GLOBal Infrastructures for Supporting Biodiversity research (GLOBIS-B) project (www.globis-b.eu) funded by the European Union Horizon 2020 Programme, grant no. 654003 (2015 - 2018).

Manifesto icons made by Freepik, Smashicons and Smartline from www.flaticon.com. Flaticon is licensed by Creative Commons, CC BY 3.0.

Photograph of the Promenade, Bari, Italy sourced from website www.italia.it, which promotes the brand Italy in the tourism sector.