

8th International  
Conference on  
**BIG DATA**  
& Data Science for Official Statistics

**BILBAO 2024**

Informing Climate Change and  
Sustainable Development Policies  
with Integrated Data

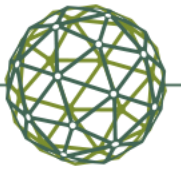
**BILBAO. SPAIN** | **10-14 JUNE 2024** | **#UNBigData2024**

# An interoperability strategy for the next generation of SEEA accounting

Ken Bagstad<sup>1</sup>, Ferdinando Villa<sup>2</sup>, Stefano Balbi<sup>2</sup>, Alessio Bulckaen<sup>2</sup>

<sup>1</sup>U.S. Geological Survey, <sup>2</sup>Basque Centre for Climate Change

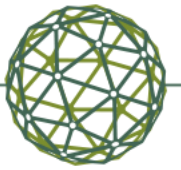




## Interoperability:

*The ability of **independently developed data\*** or tools to integrate or **work together with minimal effort***

A core challenge to the global scientific community

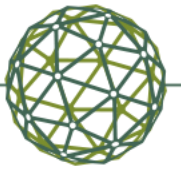


## Interoperability:

*The ability of **independently developed data\*** or tools to integrate or **work together with minimal effort***

A core challenge to the global scientific community

\*for use in **computational pipelines** – models & workflows should support interoperability too



# Types of interoperability

## Syntactic interoperability

Use of **compatible data formats** and **communication protocols**.

Low bar, more limited advantages.

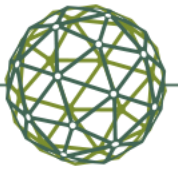


## Semantic interoperability

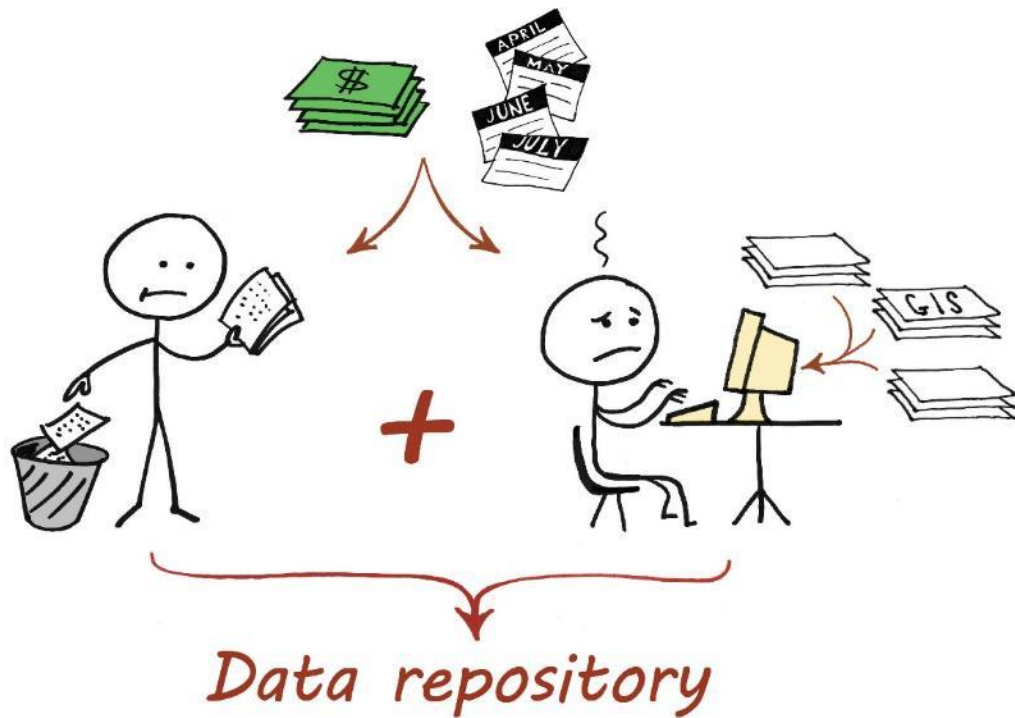
Data transfers where a **receiving system can understand the meaning of exchanged data**, reusing it appropriately.

Higher bar, greater potential for automation & data/model reuse.

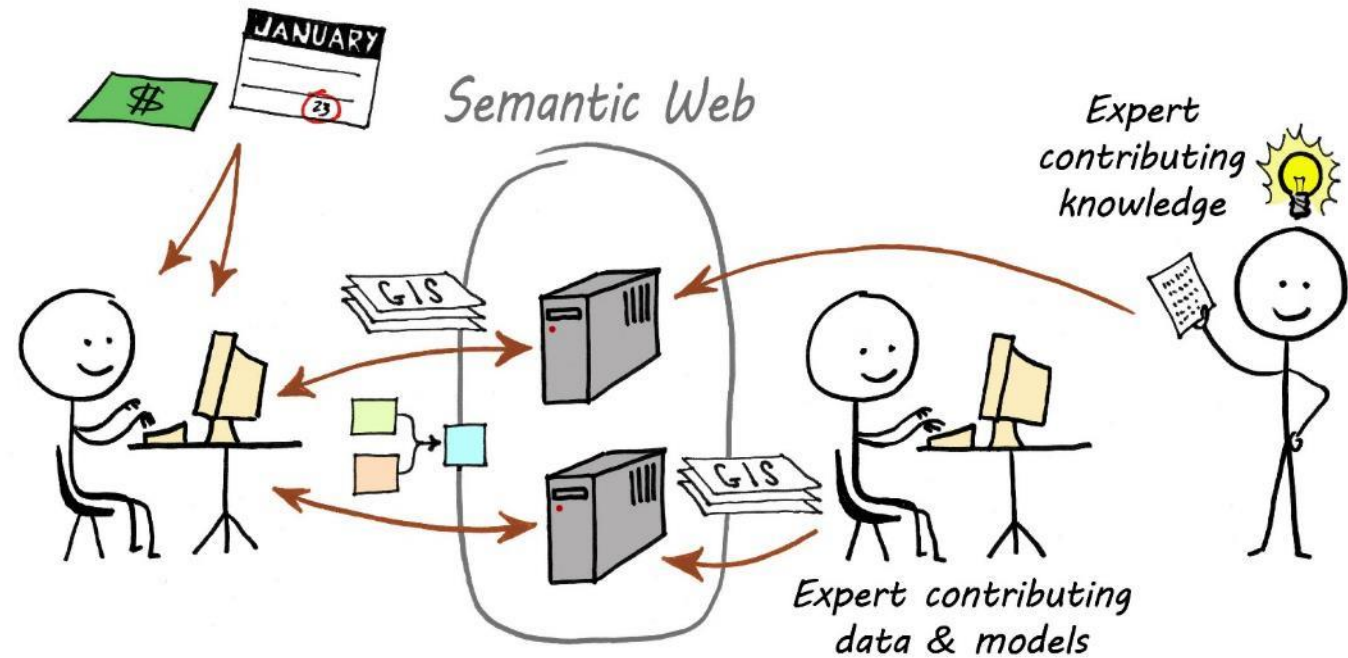
Heiler 1995



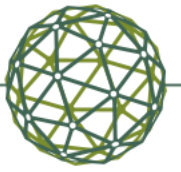
# Why? Faster, better integrative science



Status quo

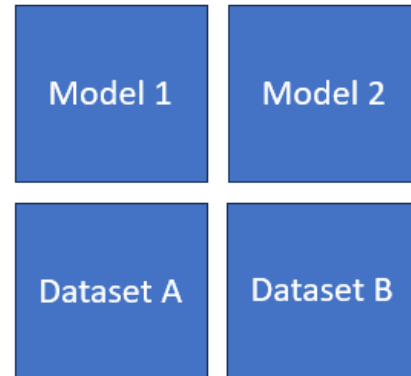
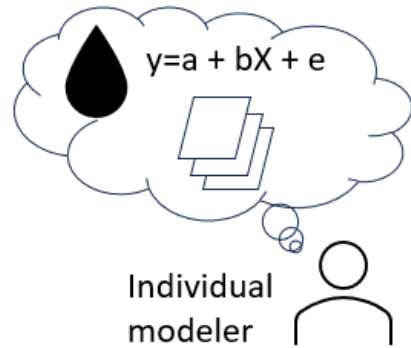


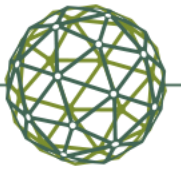
Linked, web-based collaborative modeling



# Why? Faster, better integrative science

Local scientific resources (data, models) familiar to the modeler

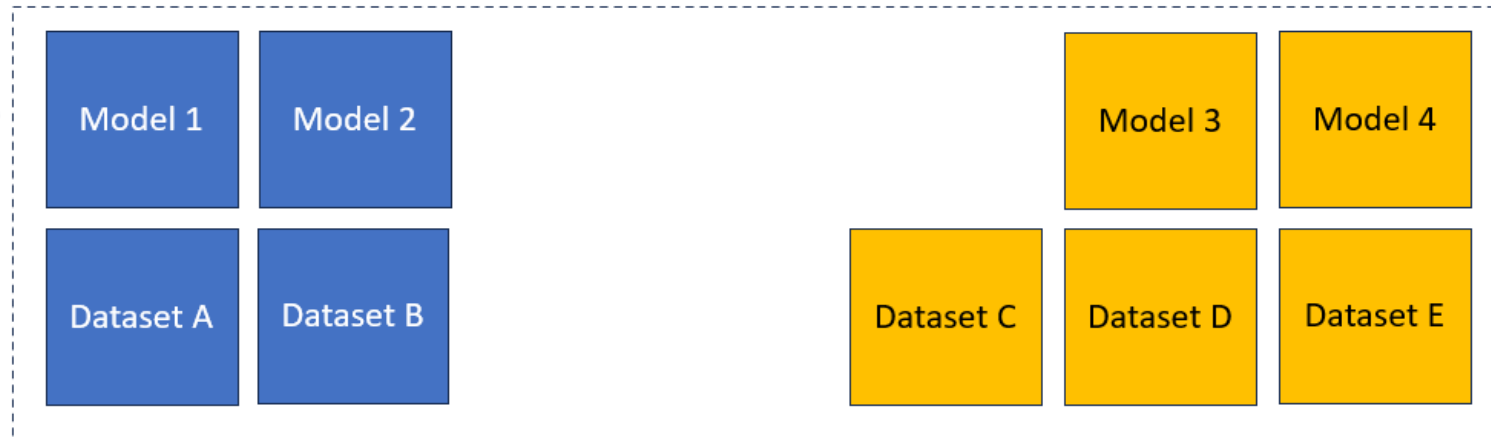
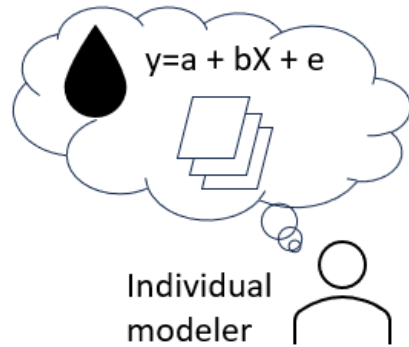


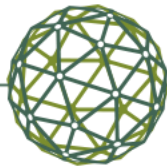


# Why? Faster, better integrative science

Local scientific resources (data, models) familiar to the modeler

Existing, interoperable scientific resources at local to global scale



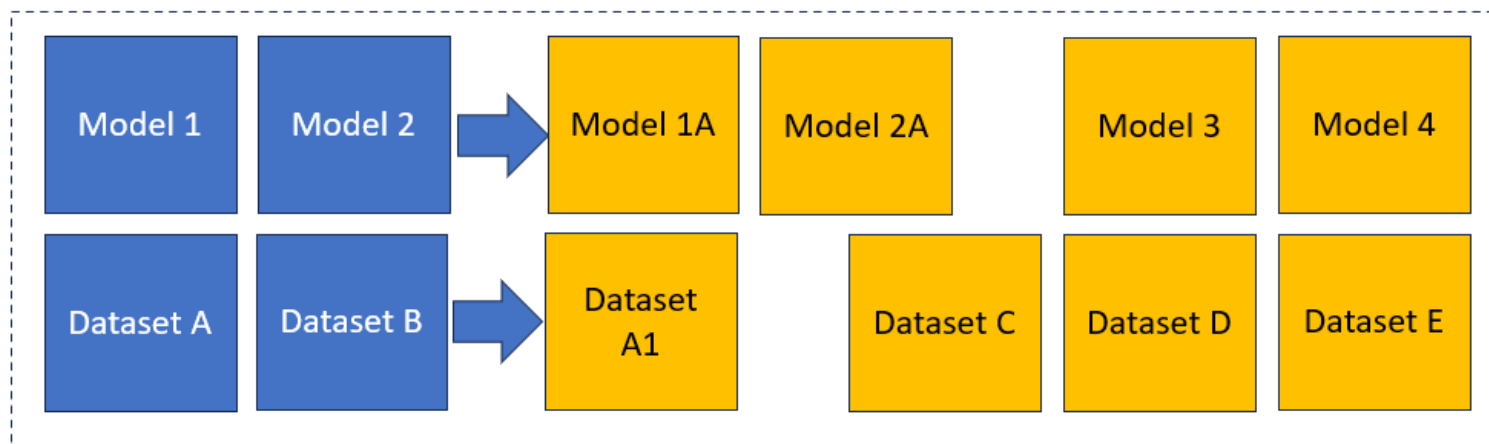
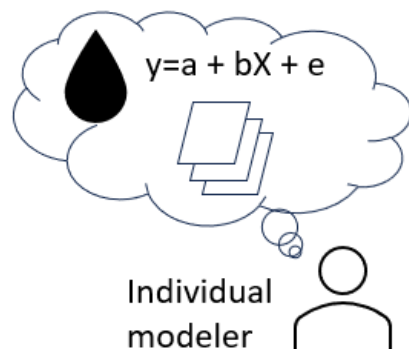


# Why? Faster, better integrative science

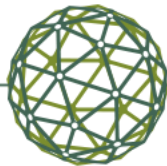
Local scientific resources (data, models) familiar to the modeler

Local resources made interoperable

Existing, interoperable scientific resources at local to global scale





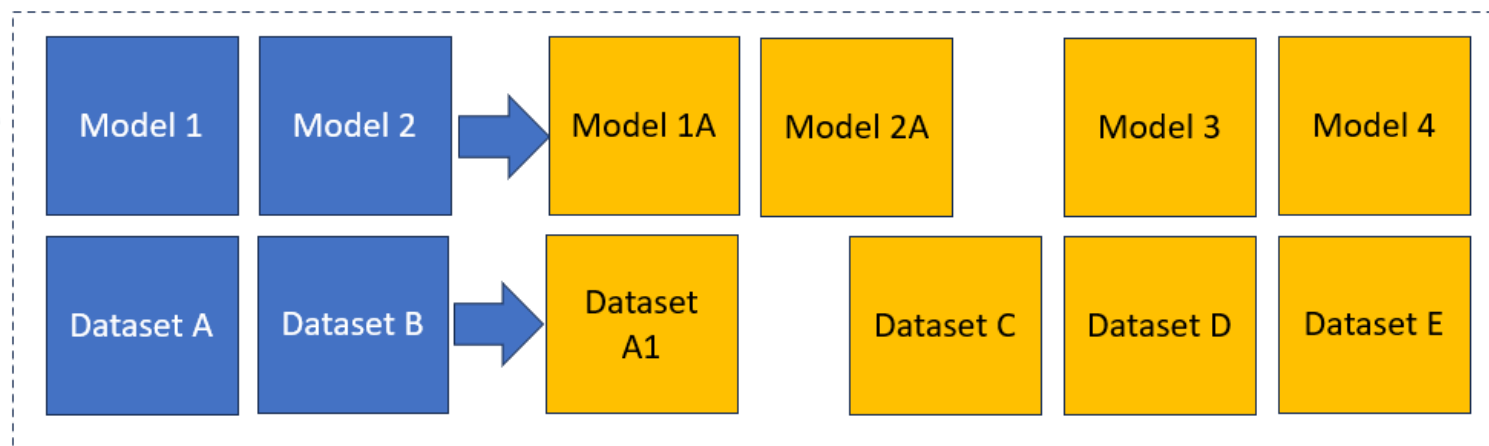
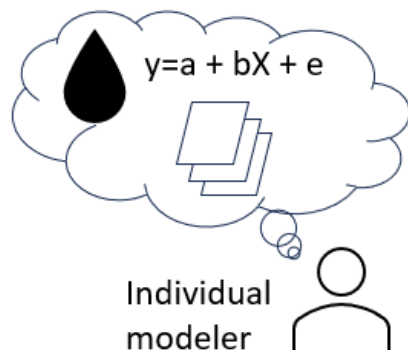


# Why? Faster, better integrative science

Local scientific resources (data, models) familiar to the modeler

Local resources made interoperable

Existing, interoperable scientific resources at local to global scale

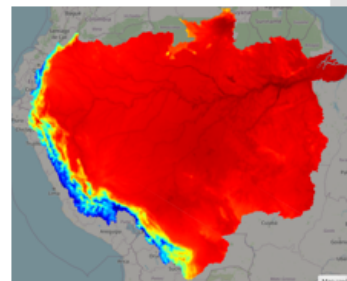


Integration through machine-supported interoperability



Results – maps, reports, provenance

- I. Introduction
- II. Methods
- III. Results
- IV. Discussion



im-data-global-climate.worldcli  
m\_21\_avg\_temp\_1970\_2000\_ja  
nuary

This processing step retrieves the contents of a data or model resource from the semantic web. Resources can be data files, data services (using protocols such as OGC or OpenAPI), or Web interfaces to more complex computations or services.

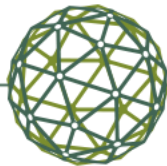
Resources are identified by a unique Uniform Resource Name (URN) used together with the scale of observation to retrieve data or trigger computation. Metadata and provenance records associated with this resource are shown below.

Title  
January average temperature, WorldClim 2.1

Originator  
WorldClim

Description  
Mean temperature (degrees C), 1970-2000 average for January at 30-second spatial resolution.

URL  
<https://www.worldclim.org/data/worldclim21.html>

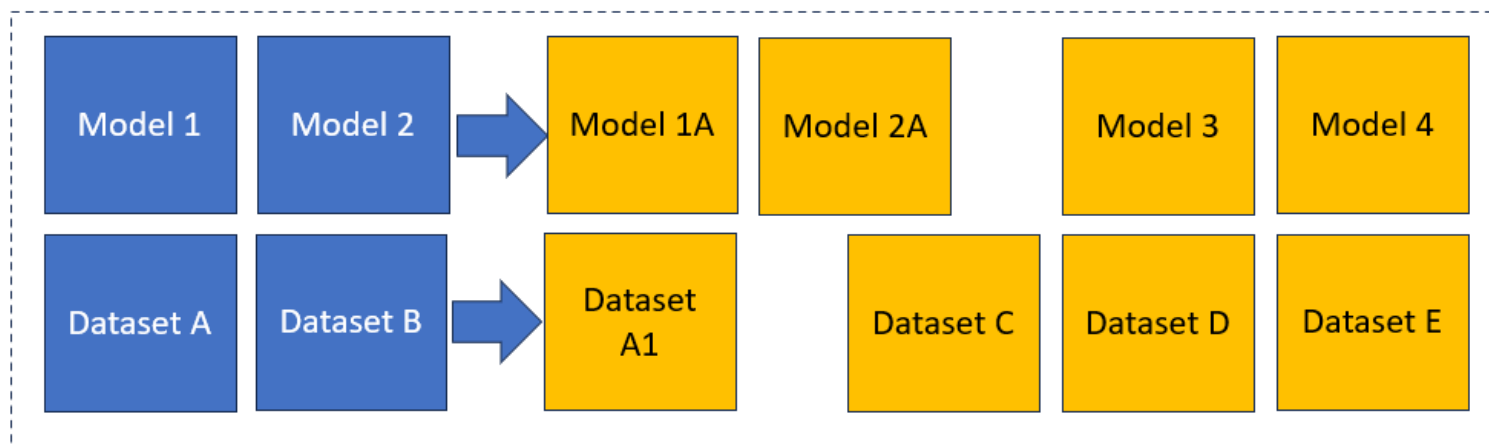
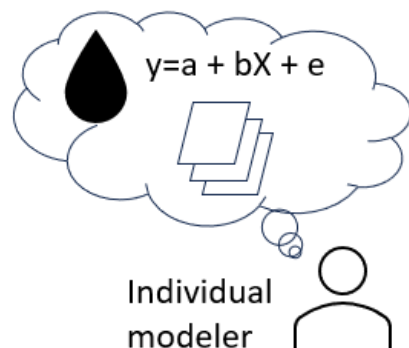


# Why? Faster, better integrative science

Local scientific resources (data, models) familiar to the modeler

Local resources made interoperable

Existing, interoperable scientific resources at local to global scale

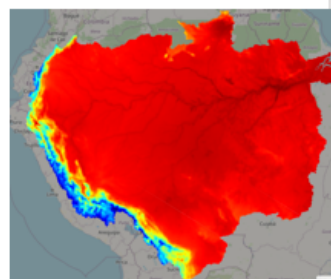


Integration through machine-supported interoperability

Iterative updating of model as new data become available

Results – maps, reports, provenance

- I. Introduction
- II. Methods
- III. Results
- IV. Discussion



im-data-global-climate.worldcli  
m\_21\_avg\_temp\_1970\_2000\_ja  
nuary

This processing step retrieves the contents of a data or model resource from the semantic web. Resources can be data files, data services (using protocols such as OGC or OpenAPI), or may interface to more complex computations or running simulations.

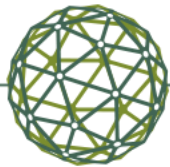
Resources are identified by a unique Uniform Resource Name (URN) and together with the scale of observation to retrieve data or trigger computation. Metadata and provenance records associated with the resource are shown below.

Title  
January average temperature, WorldClim 2.1

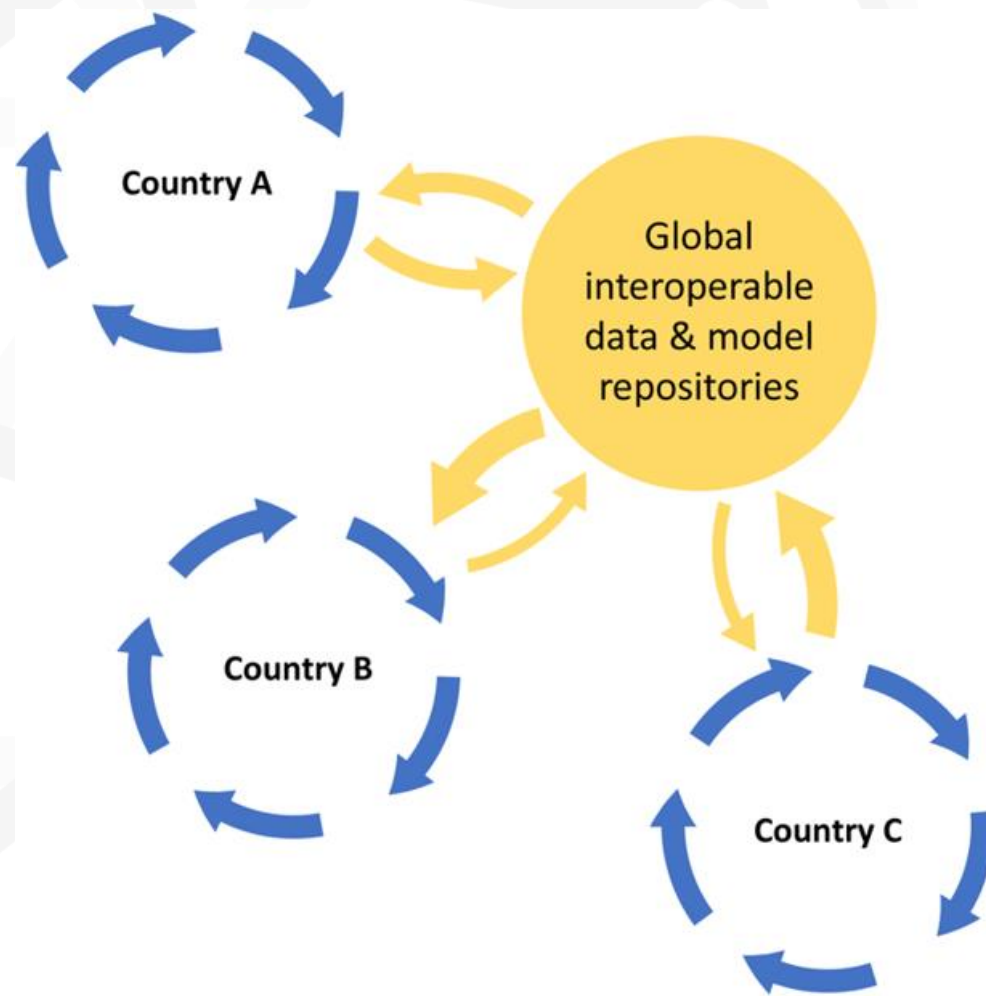
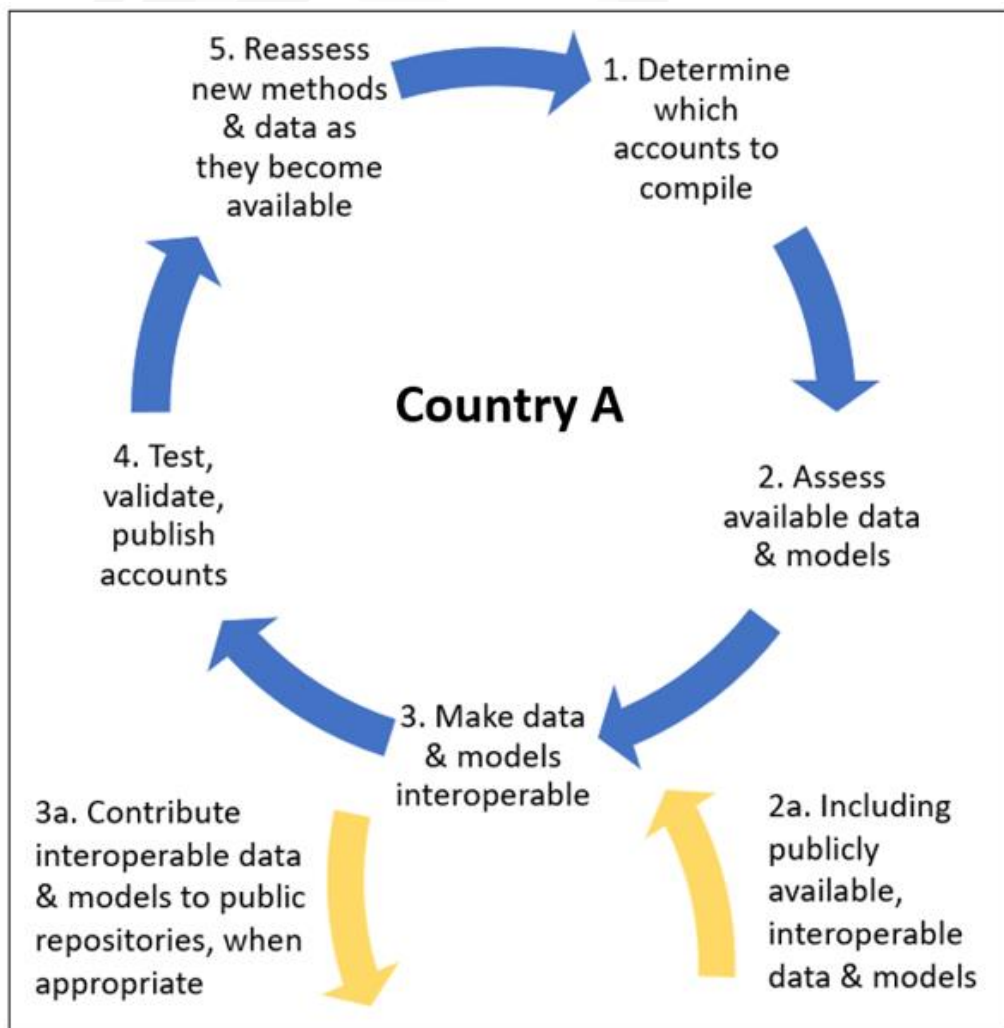
Originator  
WorldClim

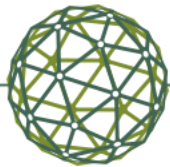
Description  
Mean temperature (degrees C), 1970-2000 average for January at 30-second spatial resolution.

URL  
<https://www.worldclim.org/data/worldclim21.html>



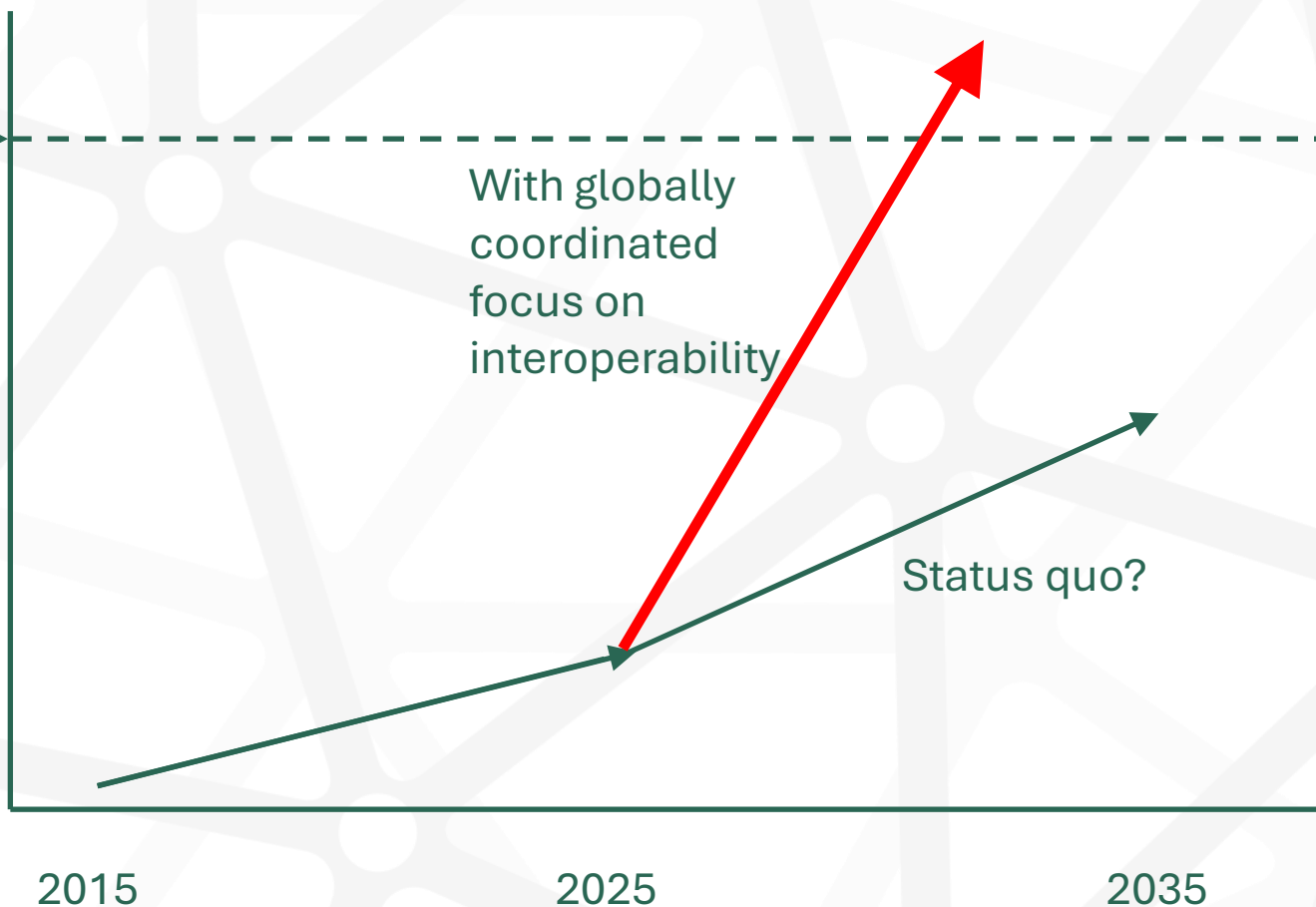
# Why? Faster implementation of ambitious global monitoring

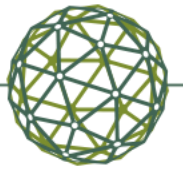




# Why? Faster implementation of ambitious global monitoring

Independent  
production, use,  
institutionalization of  
ES globally





# Major needs for interoperability

- **Consensus** & understanding of the need
- Platform-agnostic, **machine actionable** data & models
- **Semantics** (metadata descriptors) that can be navigated by AI to assemble data & models
- Reduce **hidden costs** (cloud egress costs, restrictions on private-sector use)

Balbi et al. *Environmental Evidence* (2022) 11:5  
<https://doi.org/10.1186/s13750-022-00258-y>

Environmental Evidence

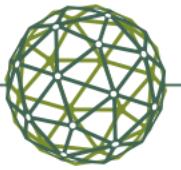
COMMENTARY

Open Access

## The global environmental agenda urgently needs a semantic web of knowledge



Stefano Balbi<sup>1,2\*</sup> , Kenneth J. Bagstad<sup>3</sup>, Ainhoa Magrach<sup>1,2</sup>, Maria Jose Sanz<sup>1,2</sup>, Naikoa Aguilar-Amuchastegui<sup>4</sup>, Carlo Giupponi<sup>5</sup> and Ferdinando Villa<sup>1,2</sup>



# Building blocks for interoperability



**1. SEMANTICS:** a flexible, shareable, easy-to-learn **language** to describe scientific observations.



**2. OPEN, LINKABLE DATA:** machine-actionable (e.g., STAC+COG/GeoServer; EPSG projected), semantically annotated data.

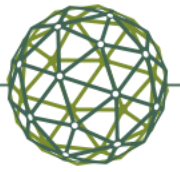


**3. OPEN, LINKABLE MODELS:** machine-actionable, “Wikipedia-like” shared, linked, semantically annotated models.



**4. GUIDELINES ON DATA/MODEL REUSE:** “guardrails” on when an AI-guided machine should reuse a particular data or model.



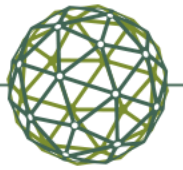


## Semantics for complex data interoperability (Leadbetter & Vodden 2015, Villa et al. 2017, Stoica & Peckham 2019, Magana et al. 2021)

1. Use **atomic concepts** to compose more complex ones
2. Distinguish the **phenomena & the property being measured** - both needed to produce a complete observation
3. Optional roles for 1) **Roles**, 2) Matrix/Realm **where measurement occurred**, 3) **measurement methods**
4. (Obvious) need for underlying **core/upper ontology** to complement domain ontologies
5. **Reuse existing** semantic resources as authorities



<https://socratic.org/questions/582d635611ef6b13ecb11fab>



# From model comparison to model interoperability

- Instead of “which model is better” (when “all models are wrong but some are useful”)
- Refocus on *when* to use different models (ecoregions, spatiotemporal scales, data availability...)
- While making them & their needed data open & interoperable
- So AI can navigate complexity & turn individual knowledge into collective intelligence

Ecosystem Services 5 (2013) e27–e39



Contents lists available at [ScienceDirect](#)

Ecosystem Services

journal homepage: [www.elsevier.com/locate/ecoser](http://www.elsevier.com/locate/ecoser)



A comparative assessment of decision-support tools for ecosystem services quantification and valuation



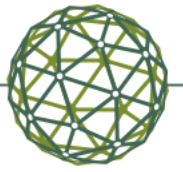
Kenneth J. Bagstad<sup>a,\*</sup>, Darius J. Semmens<sup>a</sup>, Sissel Waage<sup>b</sup>, Robert Winthrop<sup>c</sup>

<sup>a</sup> U.S. Geological Survey, Geosciences & Environmental Change Science Center, Denver, CO, USA

<sup>b</sup> BSR, San Francisco, CA, USA

<sup>c</sup> Socioeconomics Program, USDI—Bureau of Land Management, Washington, DC, USA





# Interoperability solutions must be trusted, user-friendly, equitable, community endorsed

Balbi et al. *Environmental Evidence* (2022) 11:5  
<https://doi.org/10.1186/s13750-022-00258-y>

Environmental Evidence

COMMENTARY

Open Access

## The global environmental agenda urgently needs a semantic web of knowledge



Stefano Balbi<sup>1,2\*</sup>, Kenneth J. Bagstad<sup>3</sup>, Ainhoa Magrach<sup>1,2</sup>, Maria Jose Sanz<sup>1,2</sup>, Naikoa Aguilar-Amuchastegui<sup>4</sup>, Carlo Giupponi<sup>5</sup> and Ferdinando Villa<sup>1,2</sup>



2023  
**AN INTEROPERABILITY STRATEGY FOR THE NEXT GENERATION OF SEEA ACCOUNTING**

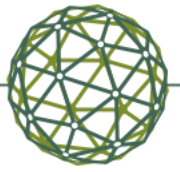


Data that powers sustainable and equitable development

# Why people are essential in data interoperability

By Steven Ramage, Jenna Slotin | August 25, 2021

<https://www.data4sdgs.org/news/why-people-are-essential-data-interoperability>



# A shared vision?



**SEEA accounts** & related indicators will be:

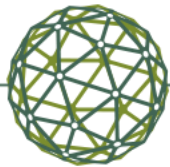
1. rapidly recompilable as new science emerges,
2. quickly produced to show the most recent trends as new annual data become available, with
3. robust international comparisons possible, while country-specific customization is still easily done.

This vision moves high-quality, meaningful **information from scientists into the hands of decision makers**, the public, and the media as quickly as possible.

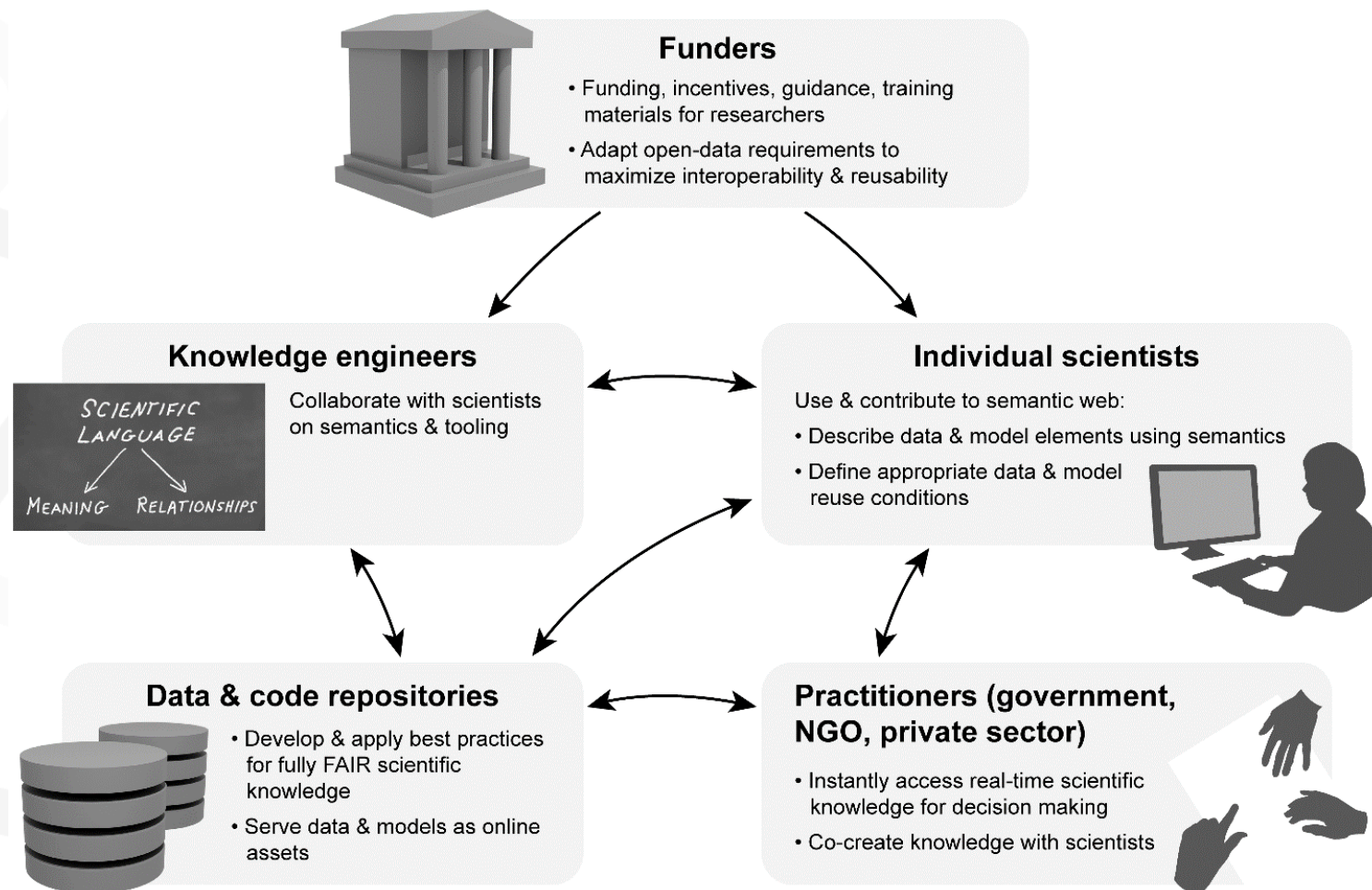
2023

**AN INTEROPERABILITY  
STRATEGY FOR THE  
NEXT GENERATION OF  
SEEA ACCOUNTING**



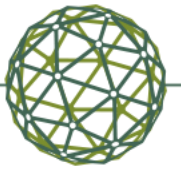


# What needs to change



Balbi et al. 2022





# What you can do

1. Learn more about interoperability & share with colleagues
2. Practice the best open-science practices you can
  1. Data & code repositories with common metadata keywords a good starting point
  2. Machine-actionable data, shared semantics, best practices for reusable code the next tier
3. Find someone in your organization savvy about this who can engage in this at the highest possible level: not everyone need be the expert

