ARIES: Artificial Intelligence for Environment & Sustainability

October 2021

Ferdinando Villa, Ken Bagstad,
Stefano Balbi & many others

USGS
science for a changing world

http://aries.integratedmodelling.org
https://integratedmodelling.org/

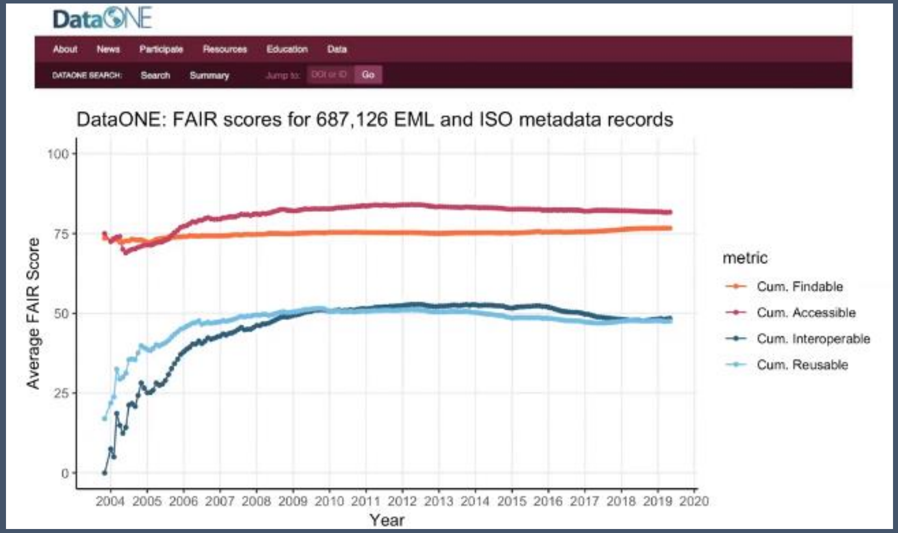
bc3
BASQUE CENTRE FOR CLIMATE CHANGE
Klima Aldaketa Ikergai

Roadmap

1. Background: The interoperability challenge & ARIES as a solution
2. ARIES Explorer tour (web interface)
3. ARIES Modeler tour (integrated development environment)
4. Getting involved
5. Q&A/discussion

I. The interoperability challenge (Borcyz & Carroll 2020)

- Individual scientists have lost the ability to read & understand all knowledge produced in their field – expensive, high-quality research fails to be reused;
- Many scientists don't care about data sharing unless they get credit; many don't feel they can trust others' data;
- Interoperability is hard to implement – requires metadata & code designed to work across different platforms;
- "Although we may not be ready now, the switch to automated research will come quickly and we should be prepared."



https://old.dataone.org/webinars/quantifying-fair-metadata-improvement-and-guidance-dataone-repository-network

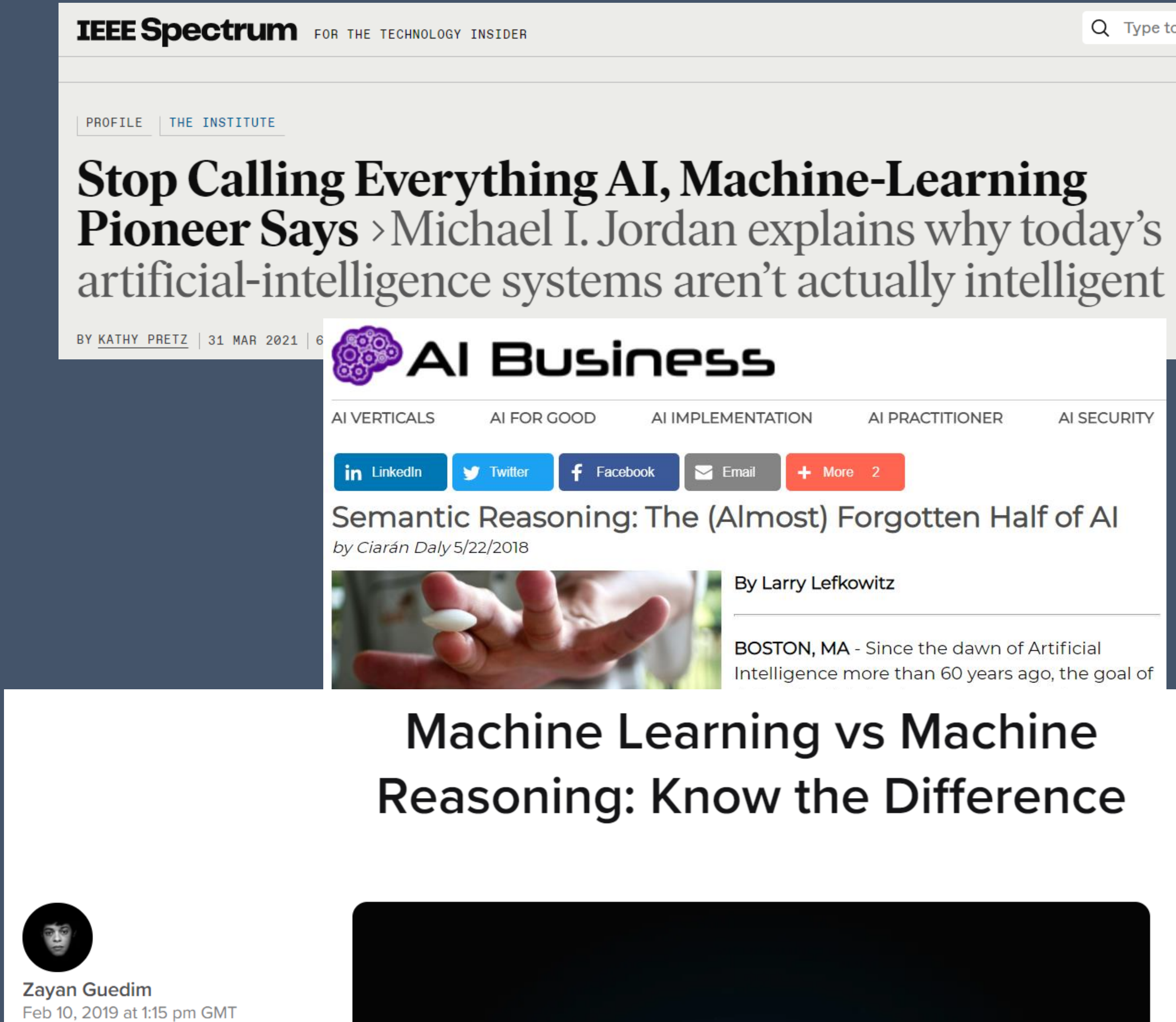
What if our open-science applications could interact with each other, automatically – i.e., through a dynamic, growing knowledge base?

Interoperability requires *common goals & standards*

- Heiler 1995:
  - Syntactic interoperability: use of compatible data formats and communication protocols. Low bar, more limited advantages.
  - Semantic interoperability: data transfers where a receiving system can properly understand the meaning of exchanged data, reusing it appropriately. Higher bar, greater potential for automation & data/model reuse.
- Interoperability in practice:
  - Go beyond current state of the practice for data & code repositories, APIs, statistical metadata (consistency & machine readability are critical)
  - Shared semantic worldviews can coordinate interaction of logically rigorous ontologies across disciplines

Machine reasoning

- "I wonder whether or when AI will crash the barrier of meaning" – Gian-Carlo Rota, 1986
- ML excels at classification & prediction, fails completely at dynamic inference
- With semantics, a knowledge base, and reasoning algorithms, computers can be taught to make decisions when presented with information



IEEE Spectrum FOR THE TECHNOLOGY INSIDER

PROFILE | THE INSTITUTE

Stop Calling Everything AI, Machine-Learning Pioneer Says › Michael I. Jordan explains why today's artificial-intelligence systems aren't actually intelligent

BY KATHY PRETZ | 31 MAR 2021

AI Business

AI VERTICALS AI FOR GOOD AI IMPLEMENTATION AI PRACTITIONER AI SECURITY

LinkedIn Twitter Facebook Email More 2

Semantic Reasoning: The (Almost) Forgotten Half of AI

by Ciarán Daly 5/22/2018

By Larry Lefkowitz

BOSTON, MA - Since the dawn of Artificial Intelligence more than 60 years ago, the goal of

Machine Learning vs Machine Reasoning: Know the Difference

Zayan Guedim

Feb 10, 2019 at 1:15 pm GMT

“The era of deep learning may come to an end”

“If somebody had written in 2011 that (deep learning) was going to be on the front page of newspapers and magazines in a few years, we would’ve been like, ‘Wow, you’re smoking something really strong,’” says Pedro Domingos.

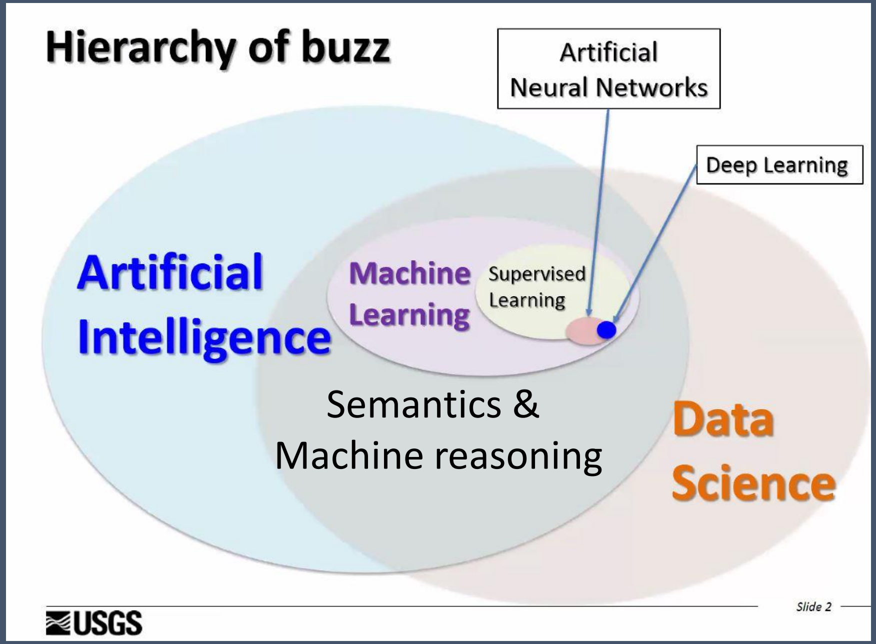
“The sudden rise and fall of different techniques has characterized AI research for a long time,” he says. “Every decade has seen a heated competition between different ideas. Then, once in a while, a switch flips, and everyone in the community converges on a specific one.”

Trendy AI topics by decade

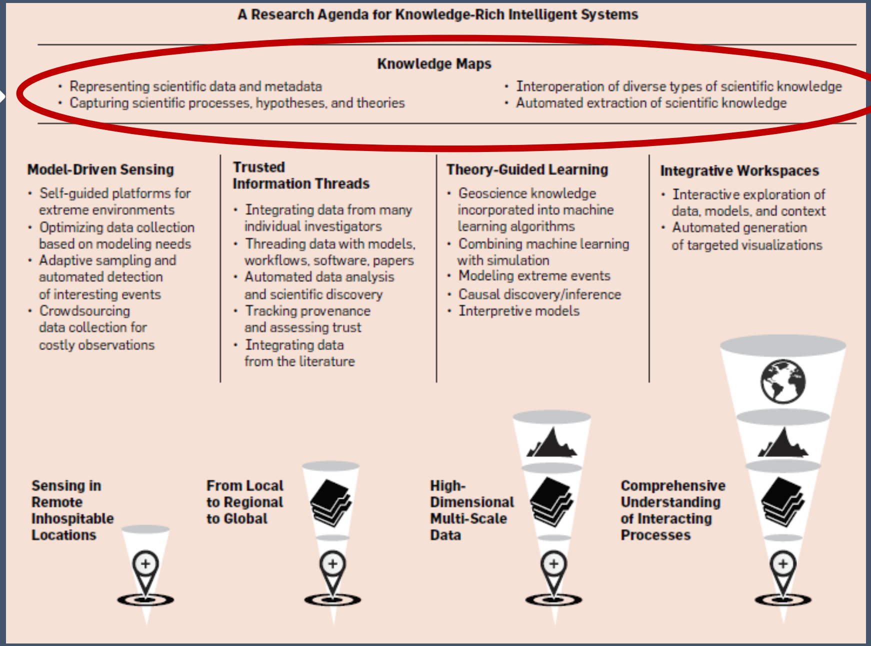
1950s & 1960s: Neural networks
1970s: Symbolic approaches
1980s: Knowledge-based systems
1990s: Bayesian networks
2000s: Support vector machines
2010s: Neural networks
2020s: ?

https://www.technologyreview.com/s/612768/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/

Artificial Intelligence for Environment & Sustainability (ARIES)



Semantics→

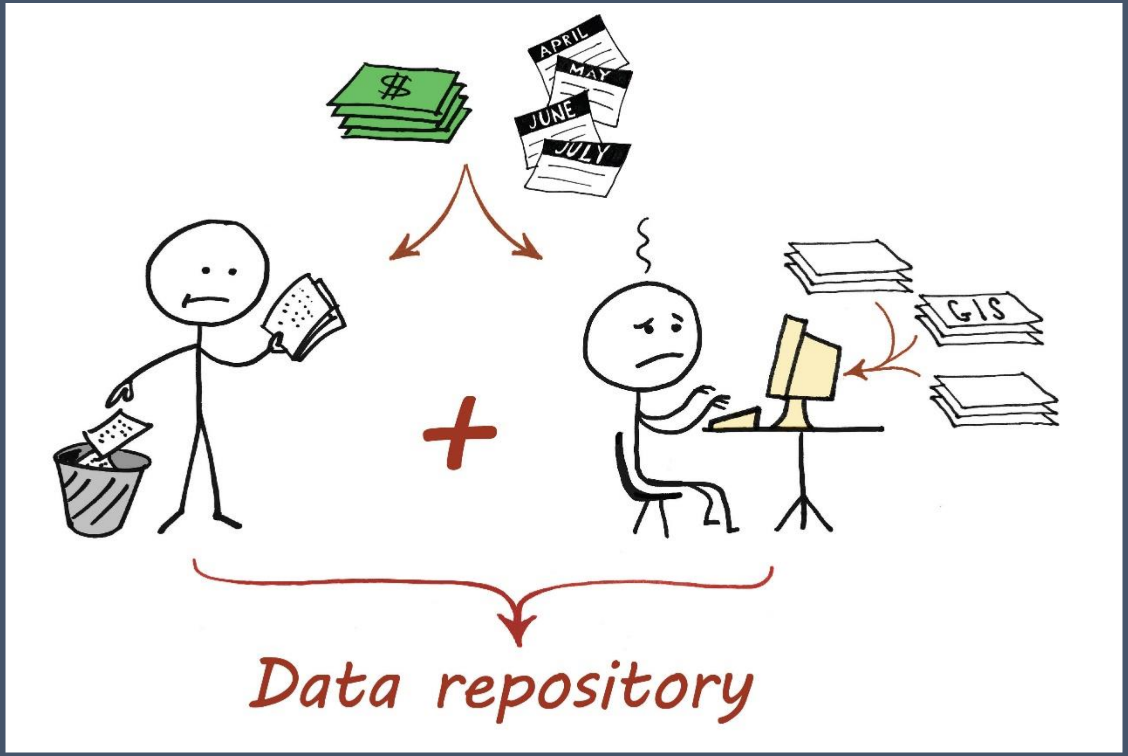


Gil et al. 2019. Intelligent systems for geosciences: An essential research agenda. Comm. ACM 62:76-84.

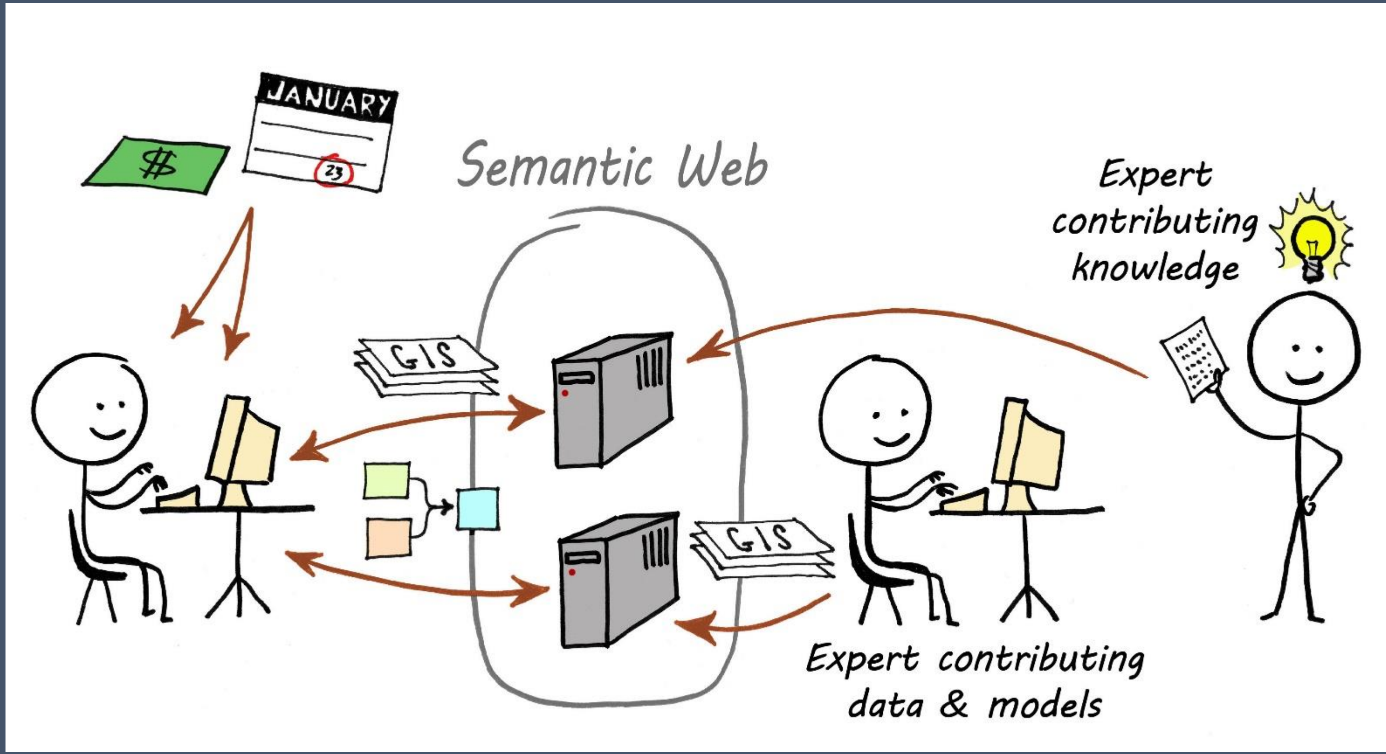
Reasoning algorithms + Decision rules + Multidisciplinary semantics + Open data & models + Open-source software

= Fast, FAIR multidisciplinary modeling

Integrated Modelling: The user perspective



Status quo



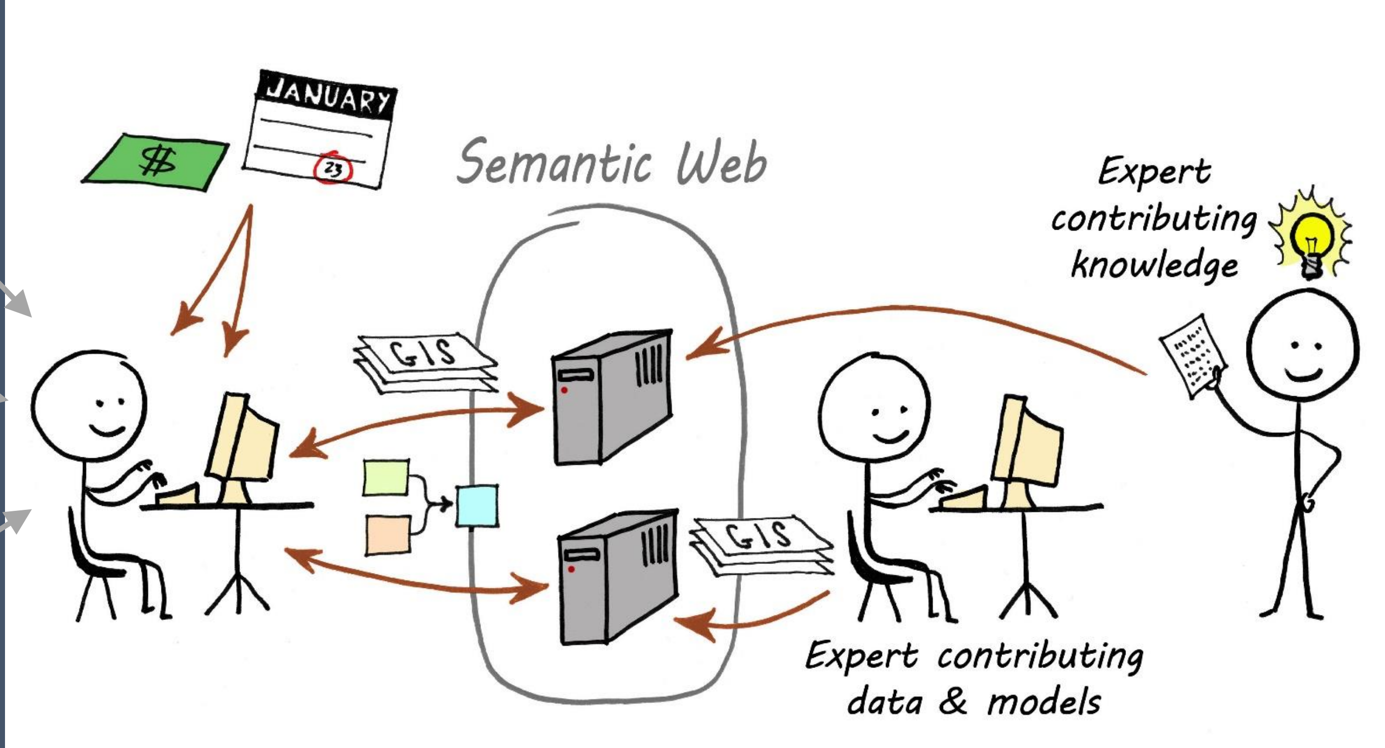
AI & semantic web-supported collaborative modeling

Who is the person on the left?

An environmental NGO

A state agency employee compiling a “state of the environment” report

A national government in the Global South



Enabling semantic interoperability

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

Developed by experts in collaboration with disciplinary scientists – typical scientist/NSO does *not* build these.

Use to accurately describe data & model elements in a consistent, machine-readable way.

2. OPEN, LINKABLE DATA: enabling access & publishing of semantically annotated data.

Put data on the web in machine-accessible formats.

Best practices already exist: no more PDFs of model parameters or zip files of spatial data.

3. OPEN, LINKABLE MODELS: open, accurate, "Wikipedia-like" sharing and linking of models.

Code models in a modular style that facilitates reuse (vs. monoliths).

Build documentation into code for automated reporting.

Specify appropriate conditions for safe reuse of your models.

k.LAB: A semantic web for science

k.LAB
Knowledge Integrated Modelling

Reactivity layer: Enables agent-based modeling, custom interactive applications & visualizations, other behaviors in response to information/conditions inspired by the original semantic web vision (Berners-Lee et al. 2001)

Semantic layer: Assigns logically consistent meaning to resources across disciplines, enabling machine reasoning

Resource layer: Nonsemantic data & model resources imported into k.LAB & assigned URNs

k.Actors (specialized language)

Semantic annotation

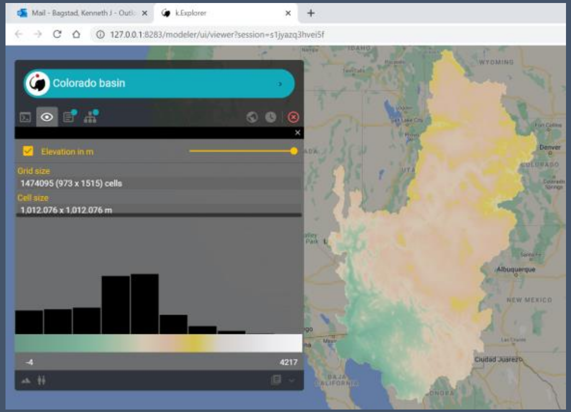
https://docs.integratedmodelling.org/technote/

Machine reasoning: How do can a machine pick the "best" data/model under which circumstances?

Initial prioritization, adjustable by advanced users; role for ML in prioritization:

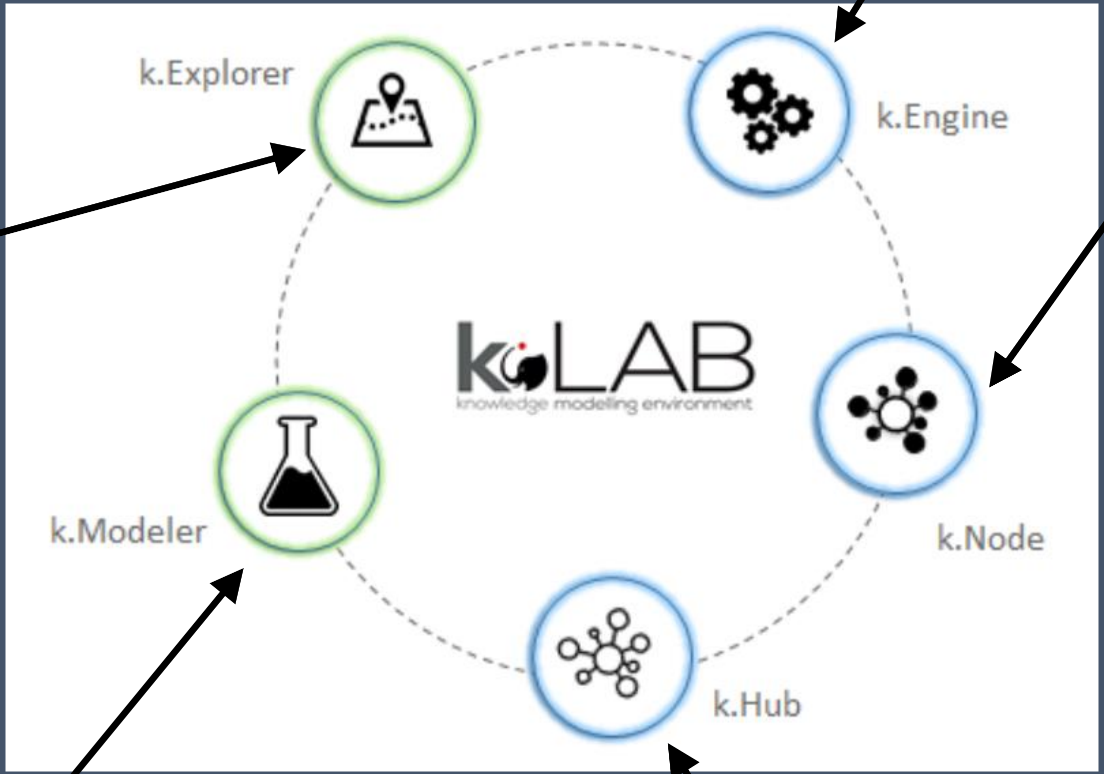
1. Lexical scope (how "close" are the data/model to the namespace & project, within k.LAB repositories)
2. Trait concordance (models with more shared attributes to the requested concept prioritized)
3. Semantic distance (concrete models prioritized over abstract ones)
4. Time specificity (closest temporal match to the user query is selected)
5. Time coverage (models with more complete temporal coverage are prioritized)
6. Space specificity (local models chosen over national, over global)
7. Space coverage (models with more complete spatial coverage are prioritized)
8. Inherency (models specified for location/scale-specific use prioritized over generalized models)
9. Evidence (data models prioritized over computed models)
10. Network remoteness (local models prioritized over those from distant networks)
11. Scale coherency (# of domains shared with the contexts)
12. Subjective concordance (user-specified metadata & weightings)

Cyberinfrastructure

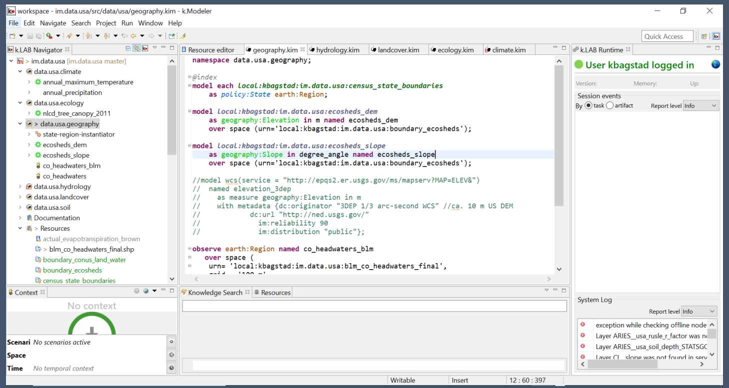


Web browser access to data & models + custom interfaces

Server implementing the modeling API & connected to the k.LAB network; enables the modeling process



Server infrastructure distributing knowledge to engines (providing knowledge for the k.LAB network – data & computed models from URNs, semantic worldviews, software components, semantic projects)

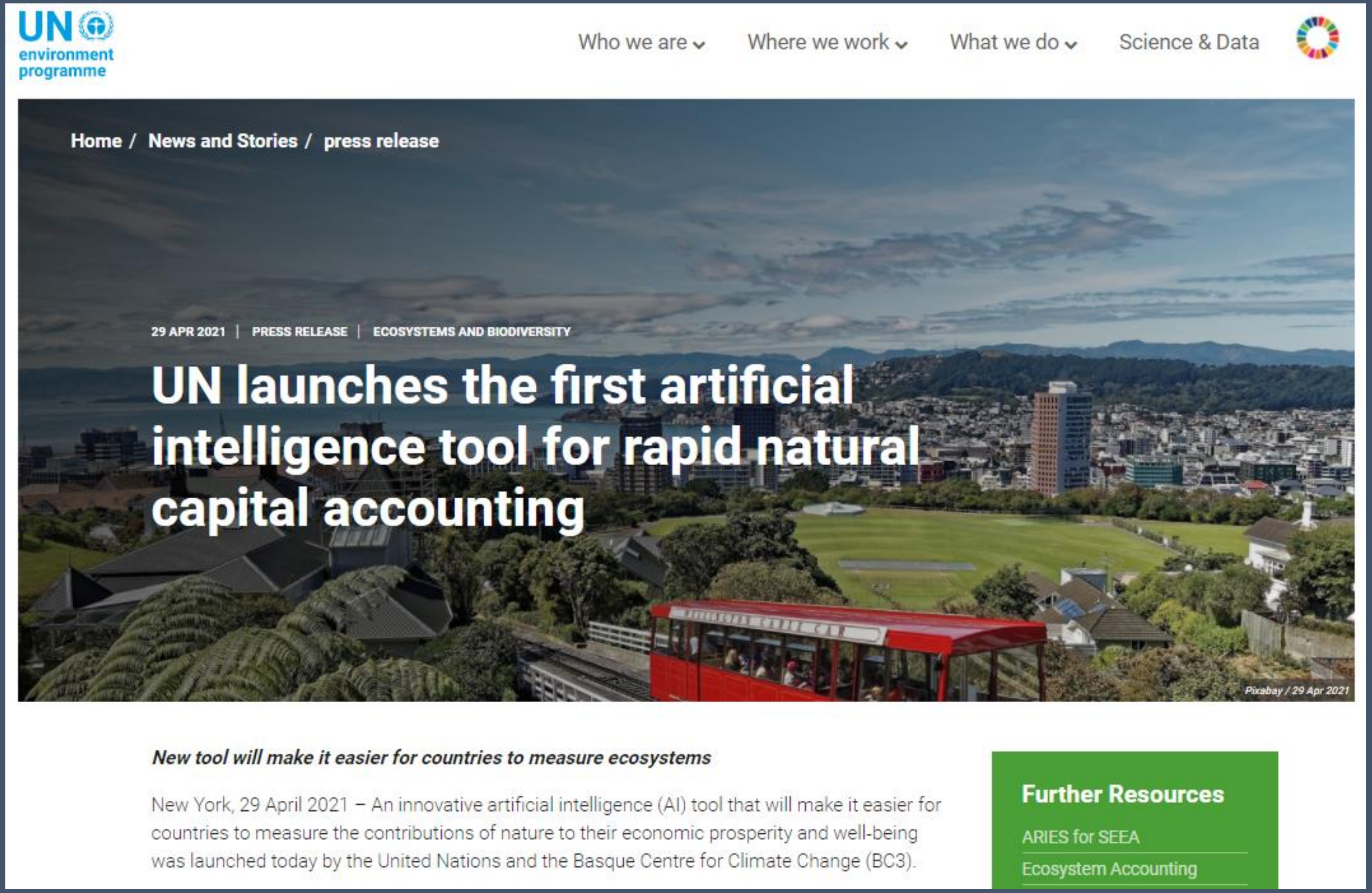


IDE for scientists to contribute & semantically annotate new data & model resources

Server coordinating k.LAB nodes, users, & engines

Example - ARIES for SEEA: Rapid, standardized, customizable environmental-economic accounting

- Global, customizable models approach enables SEEA EA compilation anywhere & improvement with local data where available
  - Fast & easy to learn
- Automate production of maps & tabular output; global modeling capacity for ecosystem extent, condition, services (Oct. 2021)
- Infrastructure for the community to share & reuse interoperable data & models



https://seea.un.org/content/aries-for-seea

ARIES for SEEA content

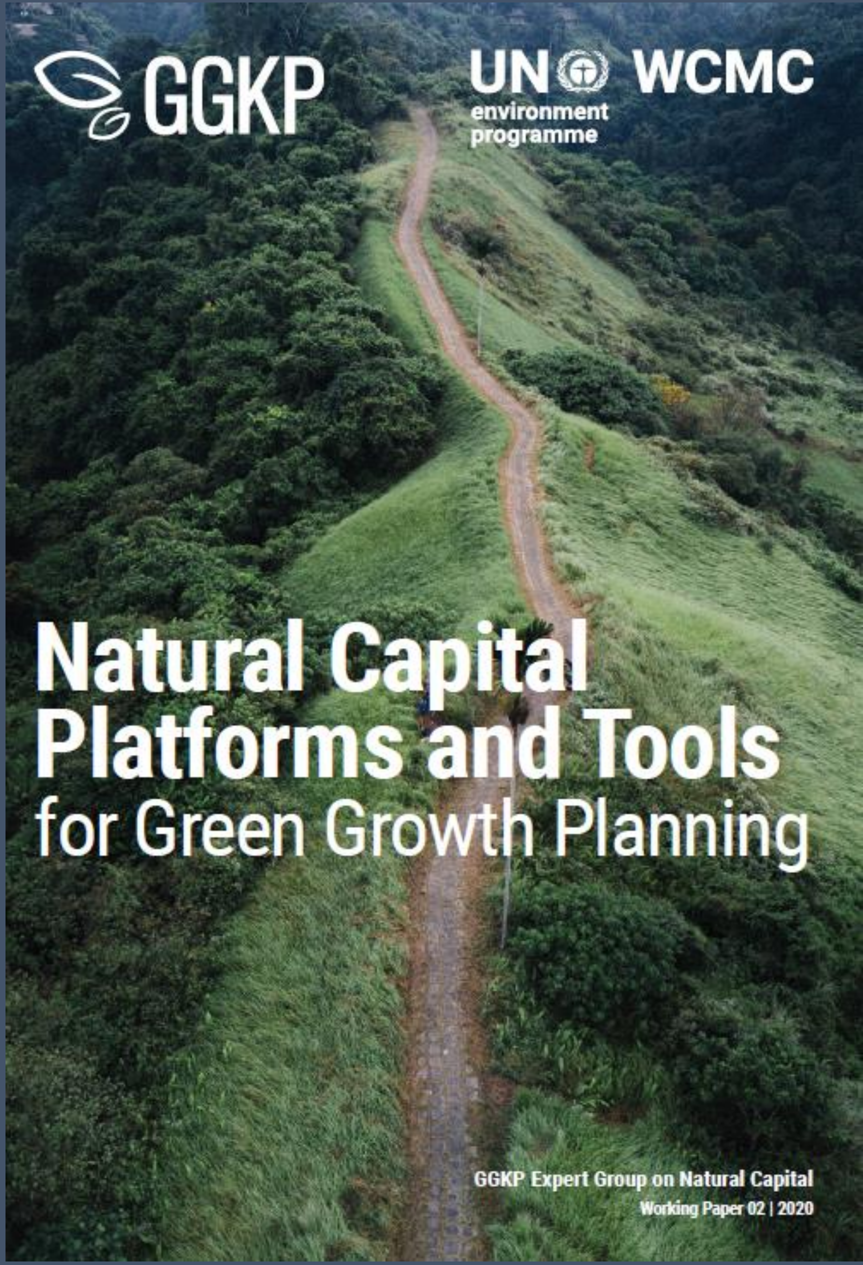
Current	Coming soon*
Ecosystem extent: 29 IUCN Level 3 ecosystem functional groups **Ecosystem condition**: 6 forest condition indicators (from GEO BON, MODIS, ESA) **Ecosystem services**: Carbon storage, crop pollination, crop provisioning, sediment regulation (physical & monetary) **Selected SDG & CBD indicators**	**Ecosystem extent**: Added IUCN Level 3 EFGs **Ecosystem condition**: Grassland & marine ecosystem condition indicators **Ecosystem services**: Nature-based tourism, water yield *Your data/models here

A shared vision for interoperable science

SEEA accounts, SDGs, CBD, & other indicators (e.g., EBVs, EOVs, EESVs) 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 from common global data, 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.



EARTH OBSERVATIONS FOR ECOSYSTEM ACCOUNTING

Articulating the vision

The global environmental agenda urgently needs a semantic web of knowledge

Stefano Balbi[1], Kenneth J. Bagstad[2], Ainhoa Magrach[1], Maria Jose Sanz[1], Naikoa Aguilar-Amuchastegui[3], Carlo Giupponi[4], Ferdinando Villa[1]

[1] Basque Centre for Climate Change (BC3) & Ikerbasque Foundation, Leioa, Spain

[2] Geosciences & Environmental Change Science Center, U.S. Geological Survey, Denver, CO, USA

[3] World Wildlife Fund, Washington, DC, USA

[4] Ca' Foscari University of Venice, Department of Economics, Venice, Italy

Complex global issues, fragmented knowledge

The global environmental agenda includes diverse internationally agreed-upon goals encompassing varied social and ecological challenges (e.g., climate change, biodiversity conservation, economic cooperation,

Technical note / k.LAB: a semantic web platform for science   Edit this Page

k.LAB: a semantic web platform for science

This document is a comprehensive, stand-alone technical introduction to the k.LAB platform, targeted to a technically savy readership. It does not substitute the k.LAB documentation and it cannot be used to learn how to use k.LAB, either from an end user or a developer perspective.

1. Introduction

Integrated modeling is a practice meant to maximize the value of scientific information by ensuring its *modularity*, *reusability*, *interoperability* and *traceability* throughout the scientific process. The open source k.LAB software, discussed

Contents

1. Introduction
2. Architecture of the k.LAB platform
2.1. The software stack
2.2. The k.LAB logical layers
2.3. Accessing the system
3. The resource layer
3.1. Resource adapters
3.2. Lifecycle of k.LAB resources
4. The semantic layer: semantic modeling
4.1. Semantic mediation and inference in support of modeling

United Nations | Department of Economic and Social Affairs | UN environment programme | ARIES | k.LAB | bc3

2021

AN INTEROPERABILITY STRATEGY FOR THE NEXT GENERATION OF SEEA ACCOUNTING

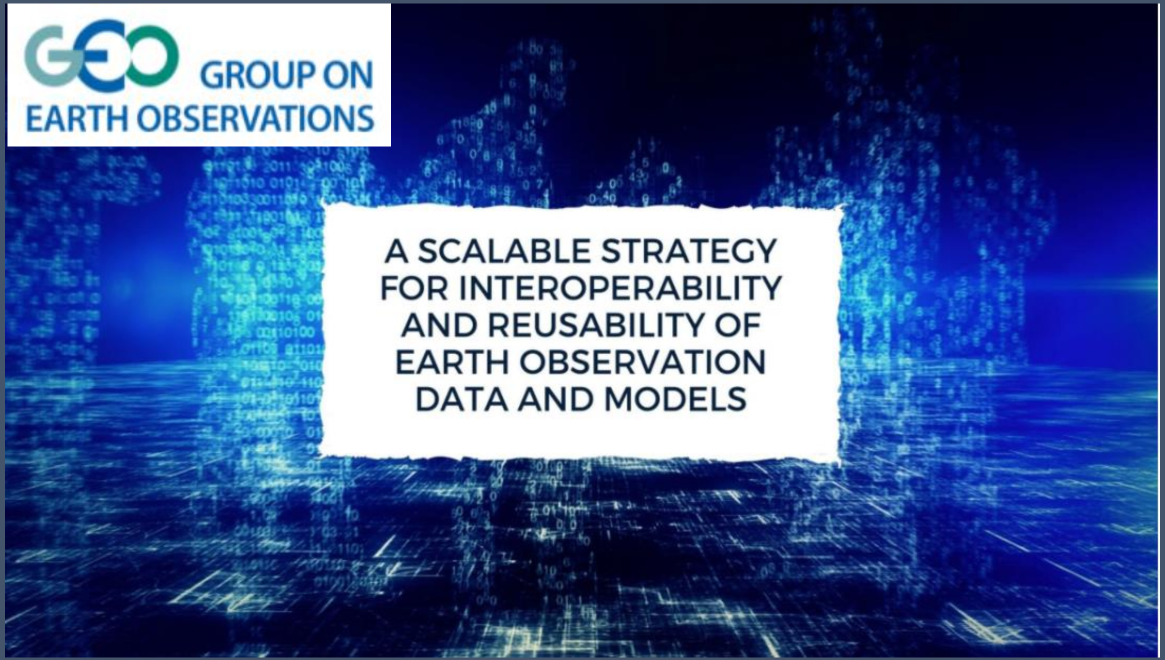
https://seea.un.org/sites/seea.un.org/files/seea_interoperability_strategy.pdf

https://docs.integratedmodelling.org/technote/

Roles of key stakeholders

- Data providers (NSOs, science agencies, academic scientists): agree on & provide data using common formats & hosting protocols (e.g., SDMX)
- Modelers (science agencies, academic scientists): use modeling practices that will make models more easily linked & documented (more modular, less monolithic); use community consensus semantics
- NSOs & other institutions (NSOs, space/mapping agencies, GEO initiatives, large academic collaborations): maintain interoperable data & models over the long term
- How can we, as a community, envision incentives that will move us beyond the status quo?

Interoperability *must address the human element*



https://www.earthobservations.org/geo_blog_obs.php?id=527

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

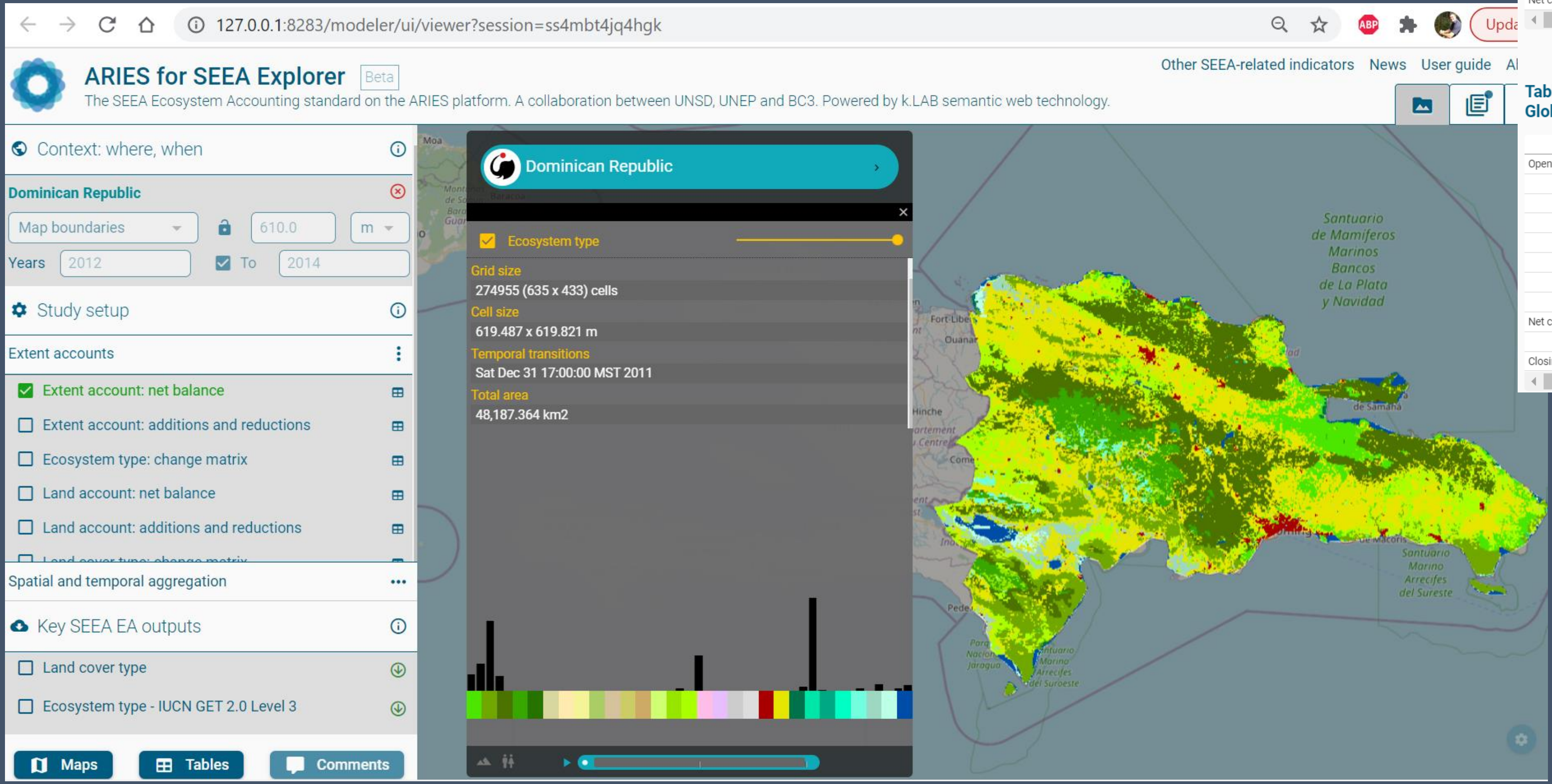
II. Web explorer tour or “How do I get my data & models into the hands of decision makers?”

Table 1. Occurring ecosystem types (selected level 3 Ecosystem Functional Groups of the IUCN Global Ecosystem Typology 2.0)

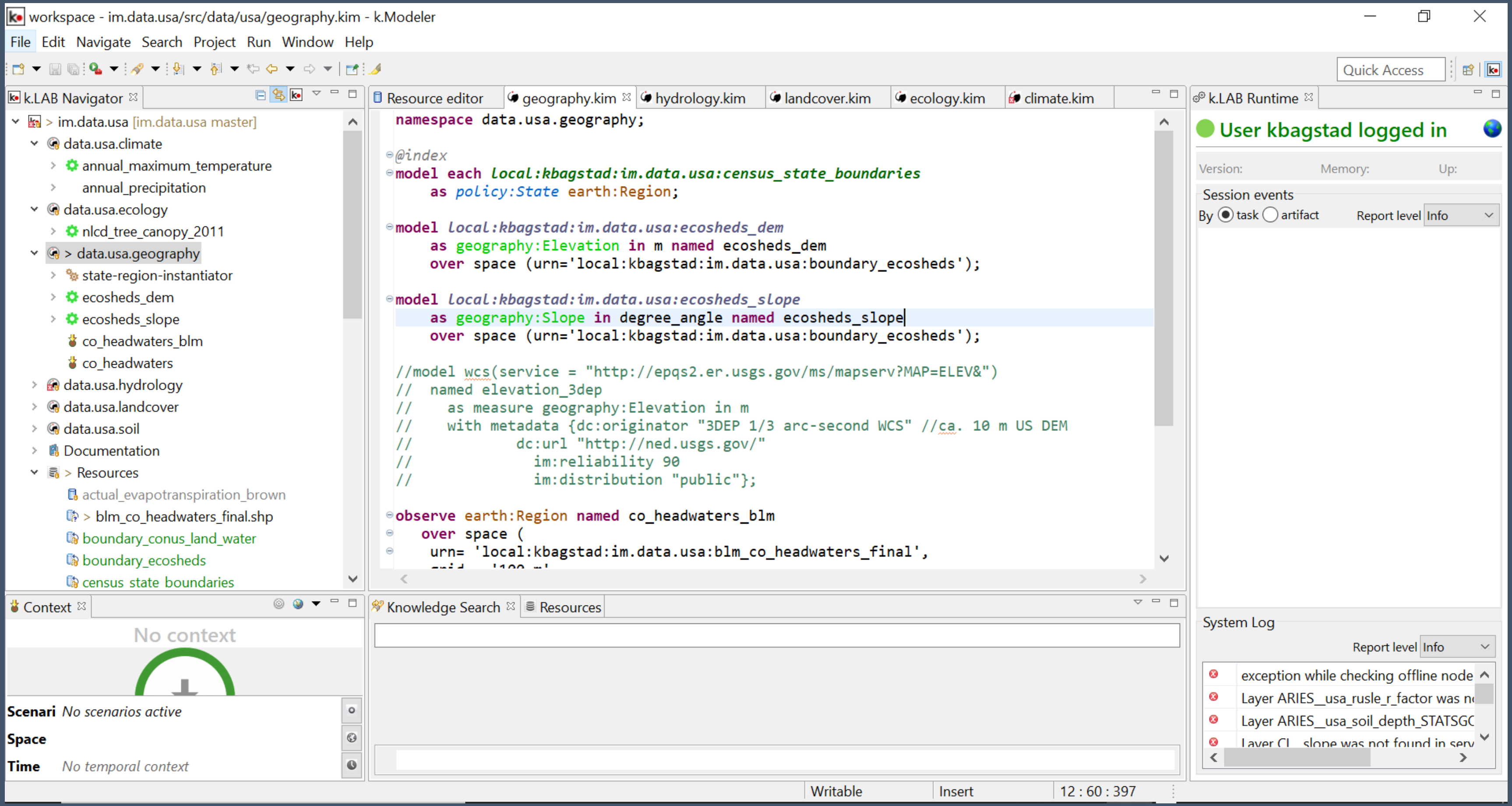
	Intertidal forest shrubland	Coastal saltmarsh reedbed	Cropland	Urban industrial ecosystem	Temperate
Extent at start of 2012 (km²)	158.25	366.39	16017.82	650.13	390.60
Extent at start of 2014 (km²)	158.25	360.81	15978.72	692.57	403.63
Net change	0.00	-5.59	-39.10	42.45	13.03

Table 2. Occurring ecosystem types (selected level 3 Ecosystem Functional Groups of the IUCN Global Ecosystem Typology 2.0)

			Intertidal forest shrubland	Coastal saltmarsh reedbed	Cropland	Urbai
Opening extent (at start of 2012)			158.25	366.39	16017.82	650.1
	Additions to extent					
		Expansions	0.00	0.00	32.39	42.4
	Reductions in extent					
		Regressions	0.00	5.59	71.49	0.00
Net change in extent			0.00	-5.59	-39.10	42.4
Closing extent (at start of 2014)			158.25	360.81	15978.72	692.

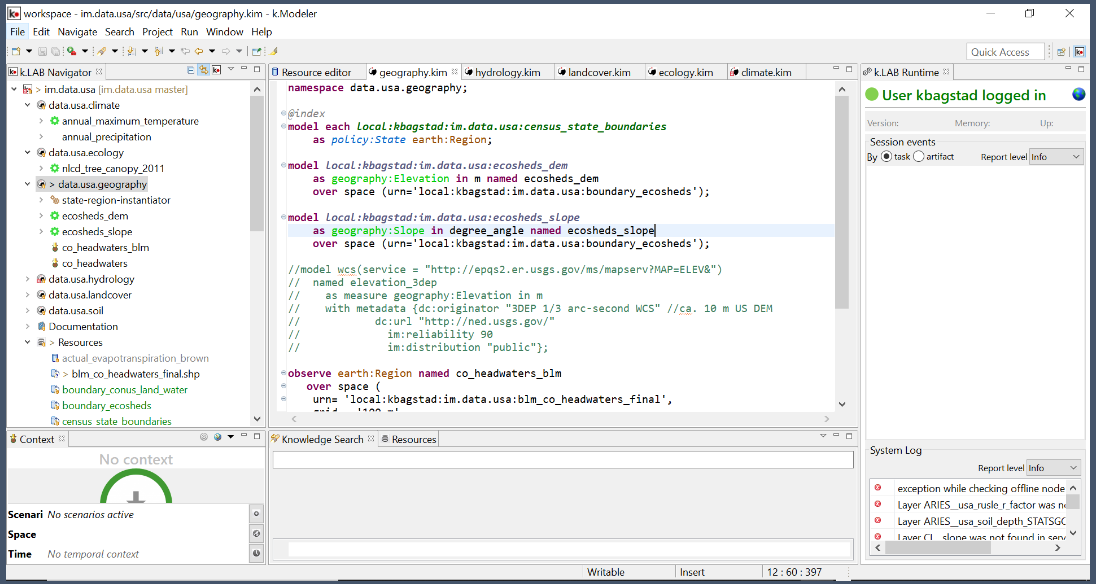


III. Integrated Development Environment (IDE) tour



Integrated Development Environment (IDE) tour

1. Projects
2. Importing resources
3. Semantic annotation of resources
4. Modeling approaches
5. Interactivity: Custom apps, tabular outputs, reporting, interactive mode



IV. Getting involved

Registering & using ARIES

- https://integratedmodelling.org/statics/pages/gettingstarted.html

Why make your science interoperable via machine reasoning?

1. Improved access to models by nontechnical stakeholders/decision makers through the ARIES Explorer (web) interface + custom applications
2. Ability to expand models' appropriate reuse, flexibility, update results as new data become available
   1. E.g., can specify appropriate conditions for model reuse (e.g., region, spatiotemporal scale) & merging of *multiple* models depending on the context of interest
3. Makes models "global yet rapidly customizable" – can run anywhere (subject to modeler-imposed reuse constraints) using common data, simultaneously substituting higher-quality local data/parameterizations where they exist. As new data become available on the ARIES network, model results can be updated, keeping results current
4. Model coupling becomes extremely easy

Why make your science interoperable via machine reasoning?

Cold Spring Harbor Laboratory

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

View current version of this article

Follow this preprint

Scientific modelling can be accessible, interoperable and user friendly: An example for pasture and livestock modelling

Alba Marquez Torres, Stefano Balbi, Ferdinando Villa

doi: https://doi.org/10.1101/2021.02.23.432363

This article is a preprint and has not been certified by peer review [what does this mean?].

Abstract | Full Text | Info/History | Metrics | Preview PDF

Abstract

This article describes the adaptation of a non-spatial model of pastureland dynamics, including vegetation life cycle, livestock management and nitrogen cycle, for use in a spatially explicit and modular modelling platform (k.LAB) dedicated to make data and models more interoperable. The aim is to deliver an existing, locally successful monolithic model, into a more modular, transparent and accessible approach to potential end users, regional managers, farmers and other stakeholders. This allows

Making data & models interoperable

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

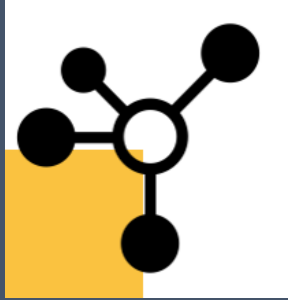
Developed by experts in collaboration with disciplinary scientists – typical scientist/NSO does *not* build these.

Use to accurately describe data & model elements in a consistent, machine-readable way.

2. OPEN, LINKABLE DATA: enabling access & publishing of semantically annotated data.

Put data on the web in machine-accessible formats.

Best practices already exist: no more PDFs of model parameters or zip files of spatial data.



3. OPEN, LINKABLE MODELS: open, accurate, "Wikipedia-like" sharing and linking of models.

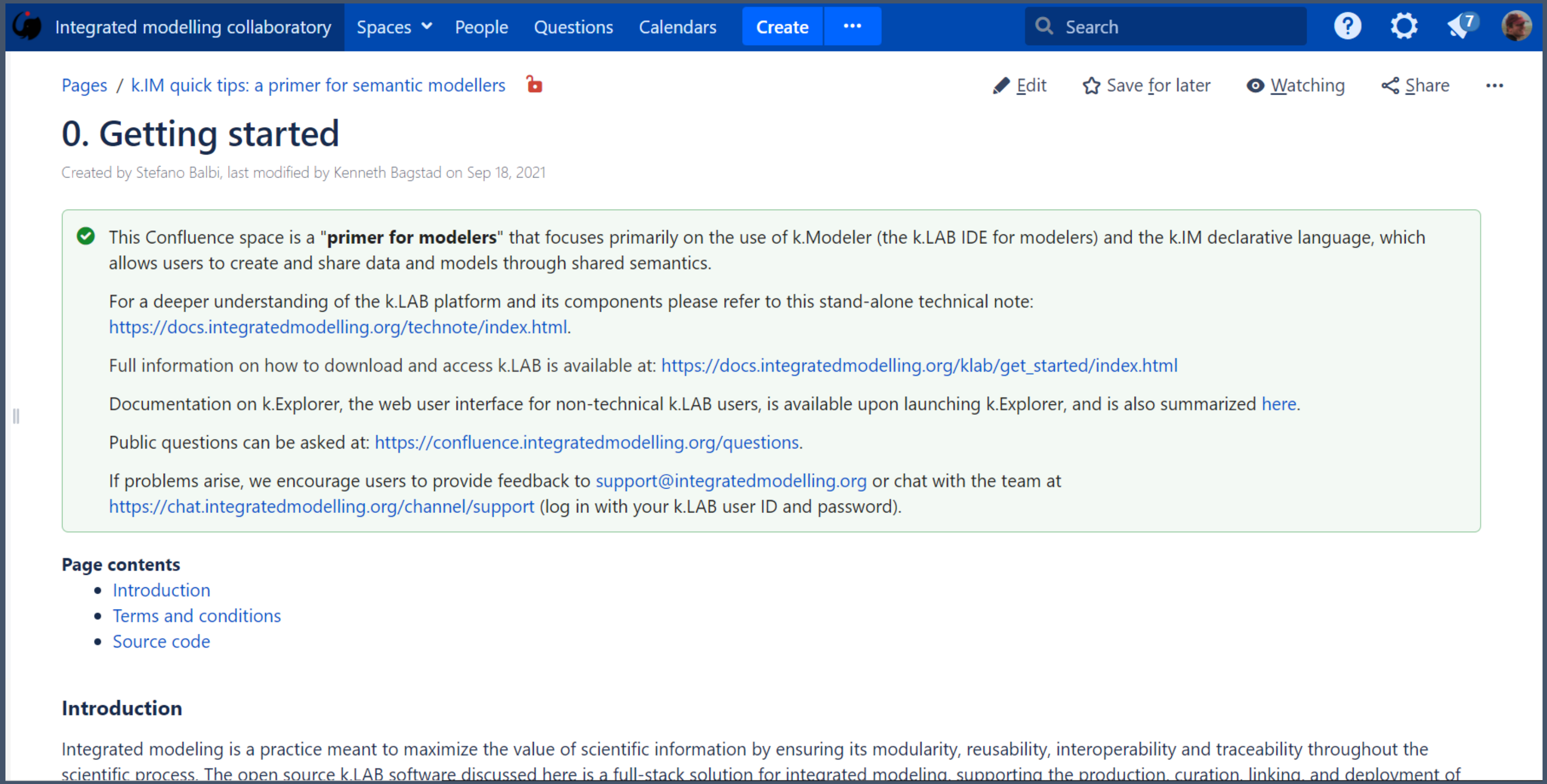
Code models in a modular style that facilitates reuse (vs. monoliths).

Build documentation into code for automated reporting.

Specify appropriate conditions for safe reuse of your models.

Tutorials & training

- Wiki space
- Questions (https://confluence.integratedmodelling.org/questions)
- Technical support (support@integratedmodelling.org)
- In-person training (https://springuniversity.bc3research.org/; hopefully 2022)
- Virtual training (coming in 2022)



Integrated modelling collaboratory Spaces People Questions Calendars Create ••• Search

Pages / k.IM quick tips: a primer for semantic modellers

Edit Save for later Watching Share

0. Getting started

Created by Stefano Balbi, last modified by Kenneth Bagstad on Sep 18, 2021

This Confluence space is a "**primer for modelers**" that focuses primarily on the use of k.Modeler (the k.LAB IDE for modelers) and the k.IM declarative language, which allows users to create and share data and models through shared semantics.

For a deeper understanding of the k.LAB platform and its components please refer to this stand-alone technical note: https://docs.integratedmodelling.org/technote/index.html.

Full information on how to download and access k.LAB is available at: https://docs.integratedmodelling.org/klab/get_started/index.html

Documentation on k.Explorer, the web user interface for non-technical k.LAB users, is available upon launching k.Explorer, and is also summarized here.

Public questions can be asked at: https://confluence.integratedmodelling.org/questions.

If problems arise, we encourage users to provide feedback to support@integratedmodelling.org or chat with the team at https://chat.integratedmodelling.org/channel/support (log in with your k.LAB user ID and password).

Page contents

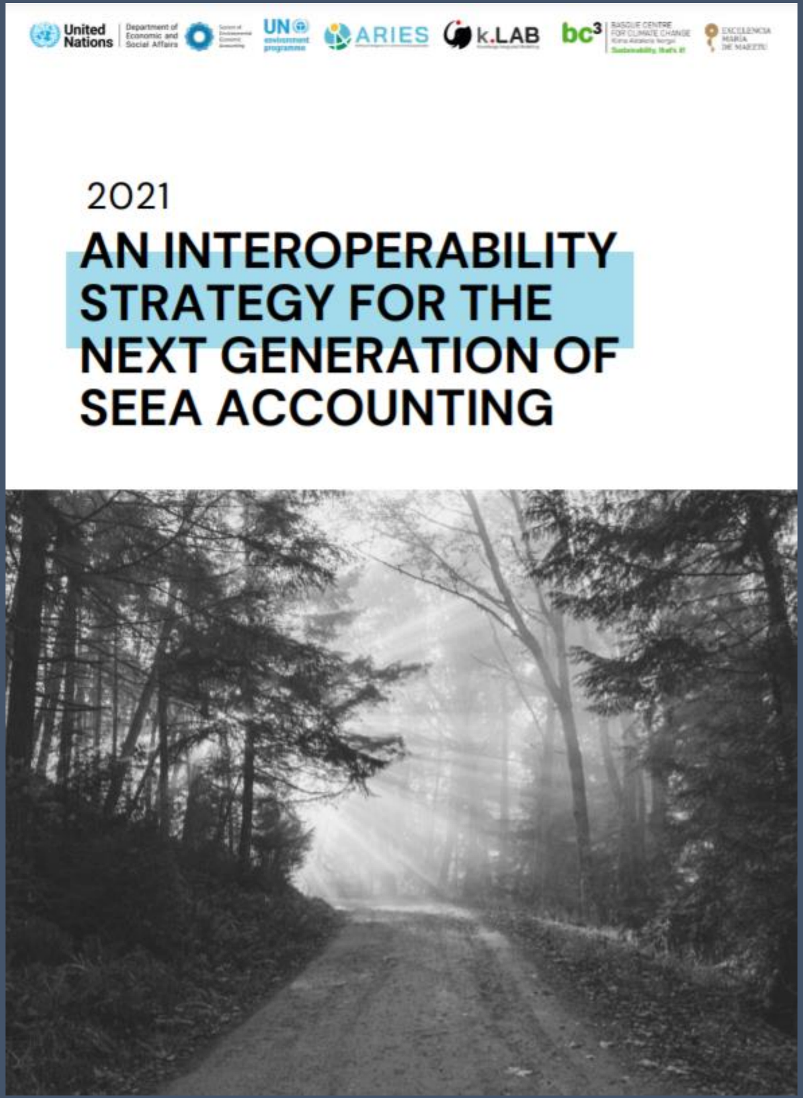
- Introduction
- Terms and conditions
- Source code

Introduction

Integrated modeling is a practice meant to maximize the value of scientific information by ensuring its modularity, reusability, interoperability and traceability throughout the scientific process. The open source k.LAB software discussed here is a full-stack solution for integrated modeling, supporting the production, curation, linking, and deployment of

https://confluence.integratedmodelling.org/display/KIM/0.+Getting+started

Learn more & support a common interoperability initiative



https://seea.un.org/sites/seea.un.org/files/seea_interoperability_strategy.pdf

Learn more

- ARIES Project: https://aries.integratedmodelling.org/
  - **Software startup guide**: https://integratedmodelling.org/statics/pages/gettingstarted.html
- Technical documentation: https://docs.integratedmodelling.org/technote/
- Interoperability strategy: https://seea.un.org/sites/seea.un.org/files/seea_interoperability_strategy.pdf
- Integrated Modelling Partnership: https://integratedmodelling.org/
- Contact kjbagstad@usgs.gov