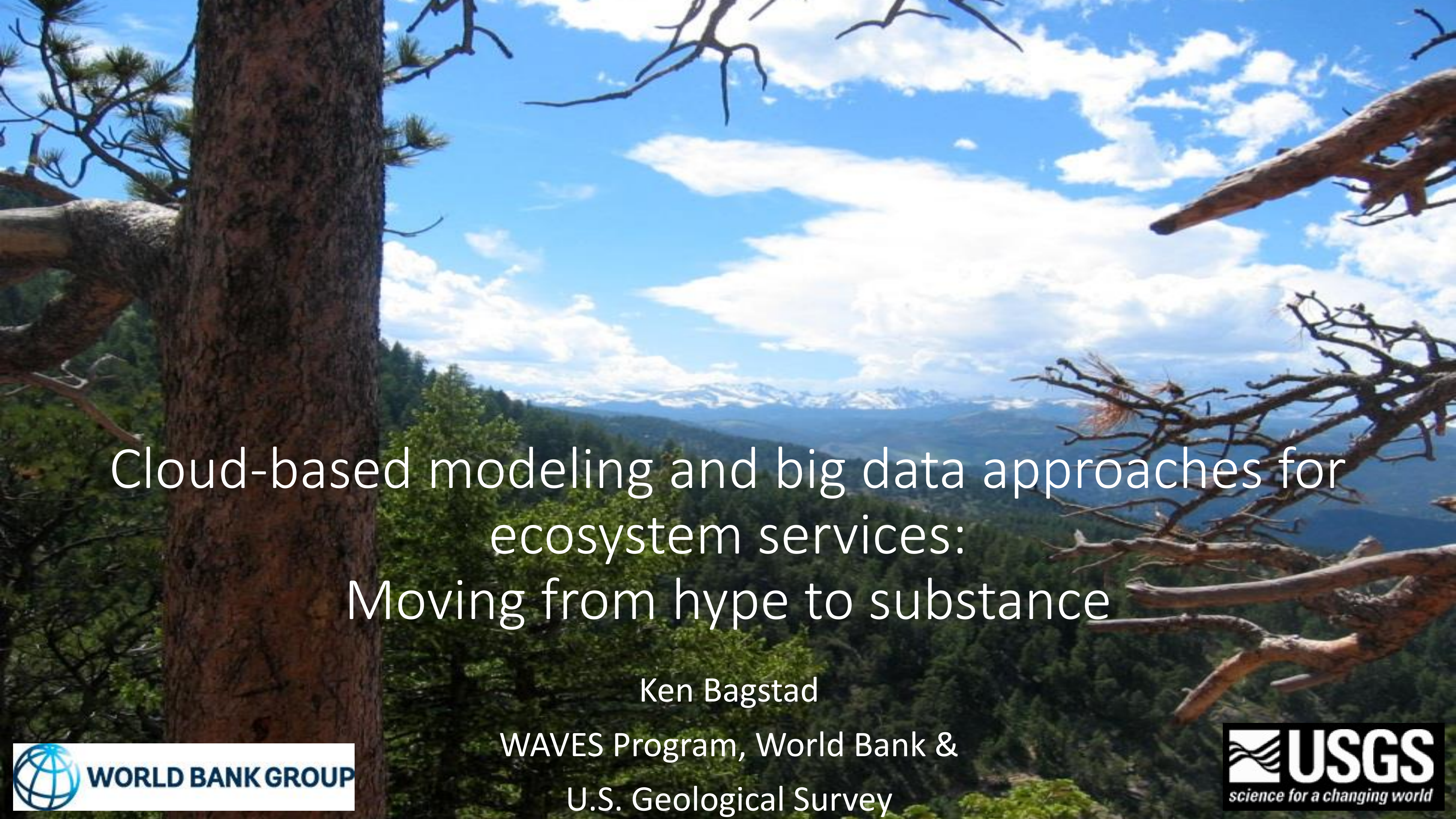


Cloud-based modeling and big data approaches for ecosystem services: Moving from hype to substance

Ken Bagstad

WAVES Program, World Bank &

U.S. Geological Survey

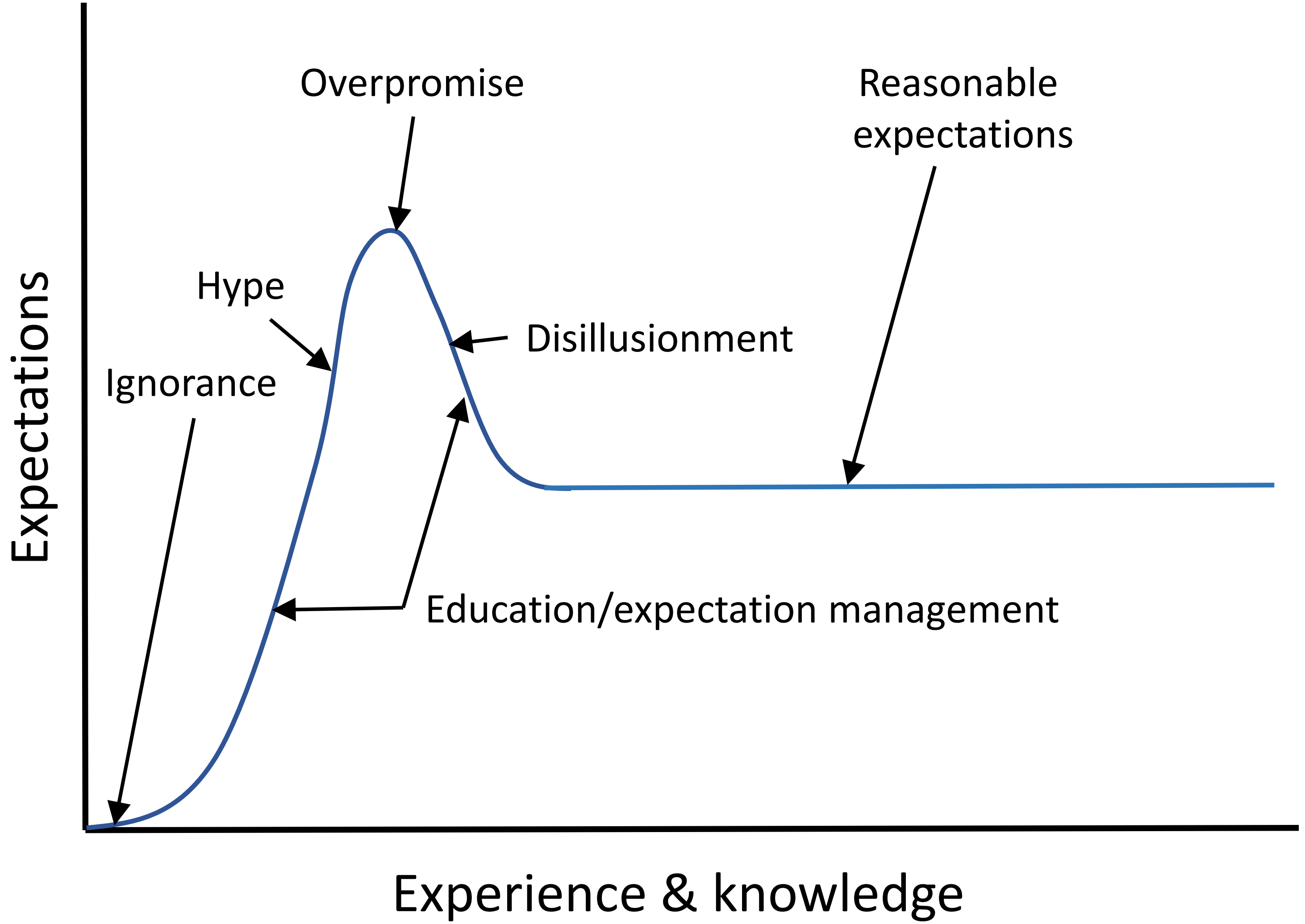
WORLD BANK GROUP

USGS science for a changing world

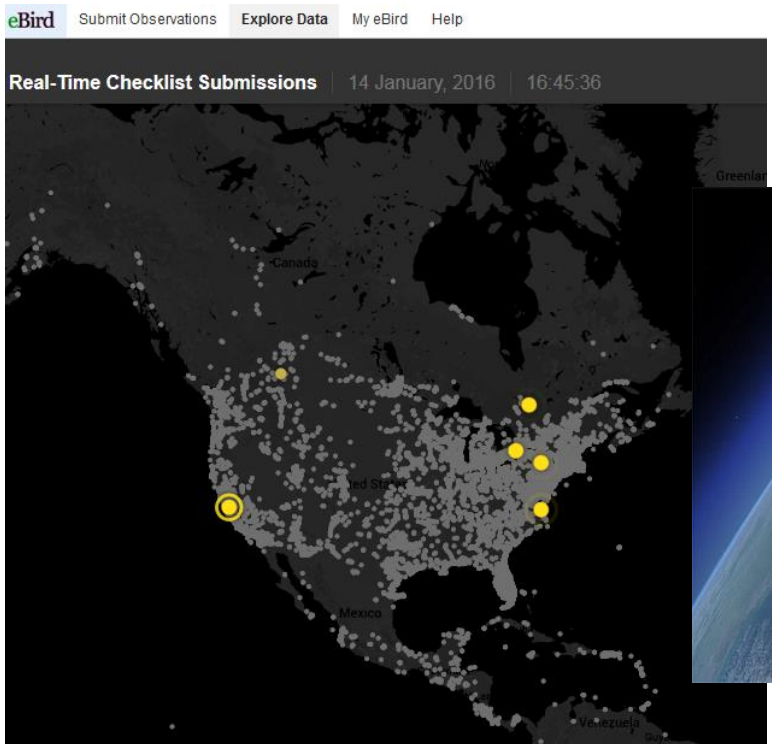
Today

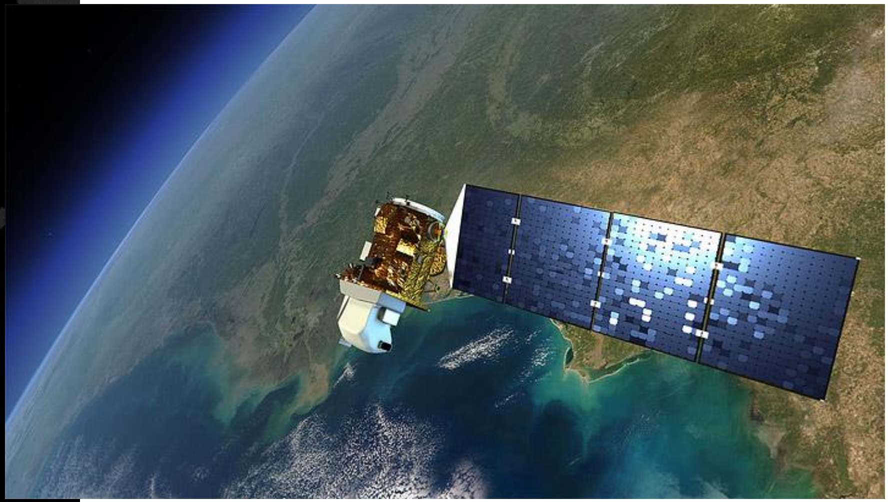
- What is big data, and why does it matter now?
- Data-intensive science
- Semantics
- Why is cloud-based semantic modeling a challenge?
- Some proposed starting points for our community

Big data: Moving from hype to reality



What is big data?

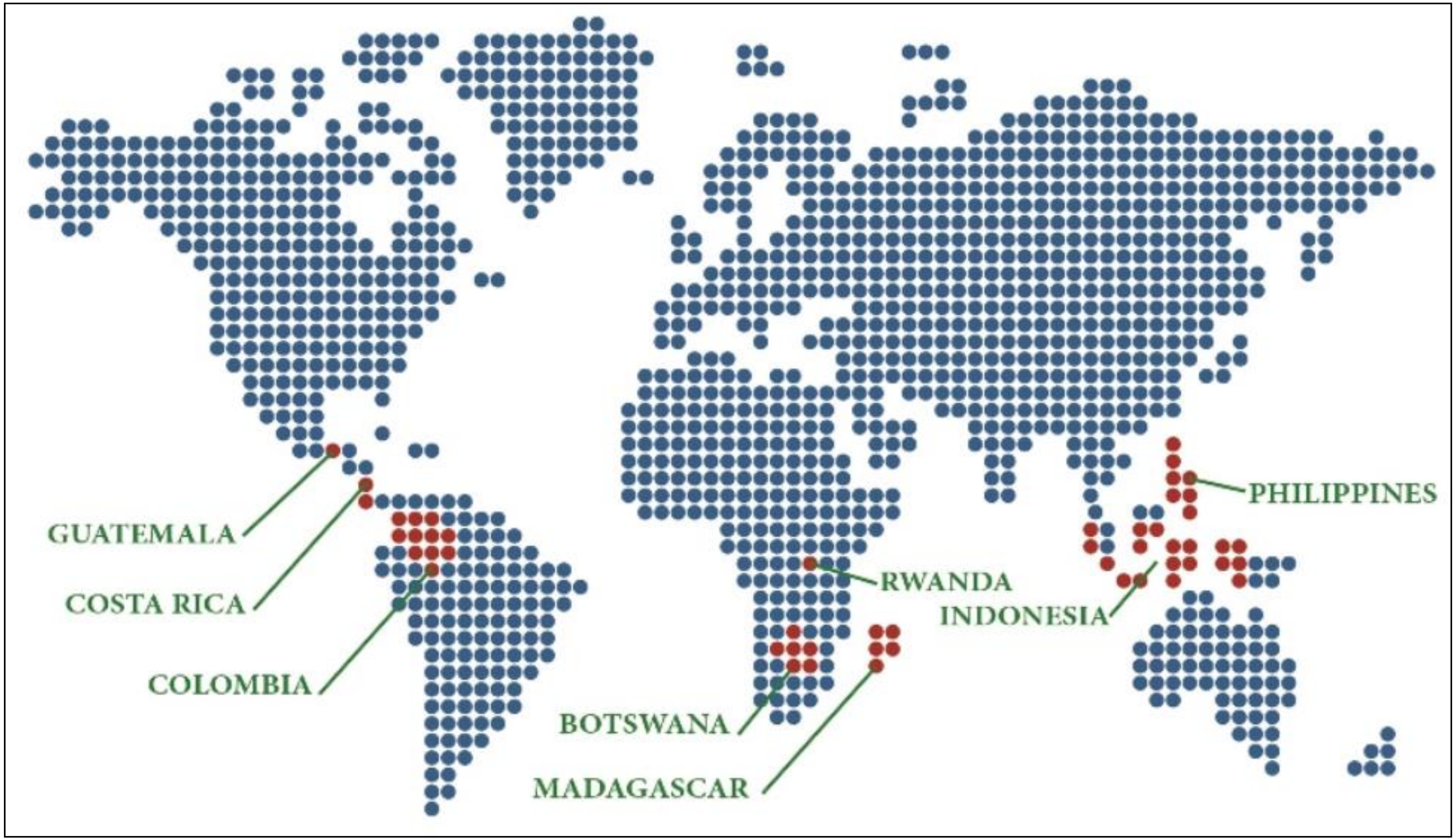




Why are we here?

1. Ecosystem services going mainstream
2. Rise of computing power & computer science
3. Proliferation of new data sources
4. We’re moving quickly, but not quickly enough

1. Mainstreaming ecosystem services

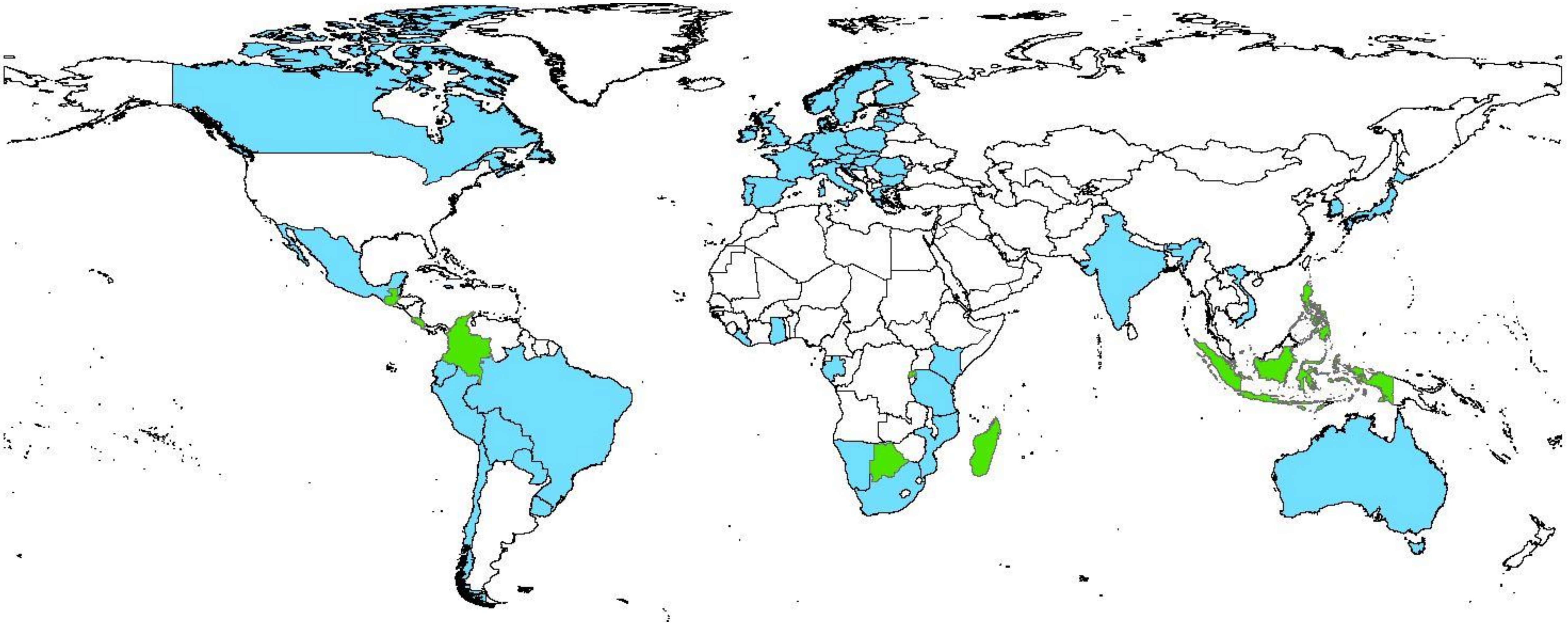


THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

WAVES
Wealth Accounting *and the* Valuation of Ecosystem Services

http://www.wavespartnership.org

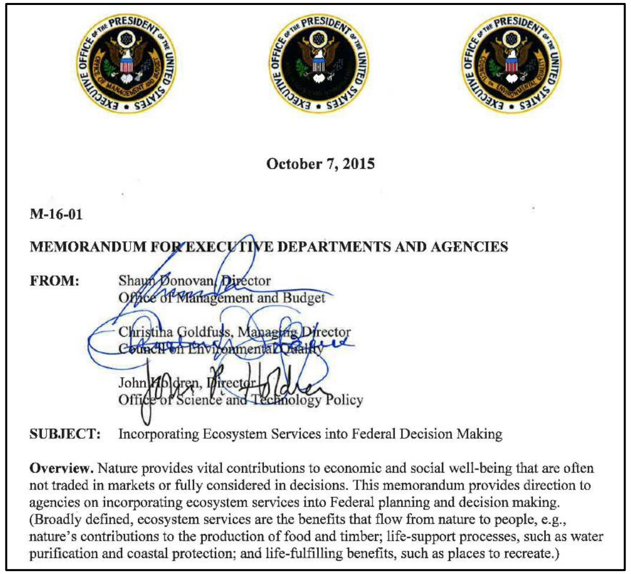
1. Mainstreaming ecosystem services



Additional countries active in natural capital accounting

1. Mainstreaming ecosystem services

Interest in: cross-jurisdictional management; evaluating tradeoffs; increasing responsivity to the public


October 7, 2015

M-16-01

MEMORANDUM FOR EXECUTIVE DEPARTMENTS AND AGENCIES

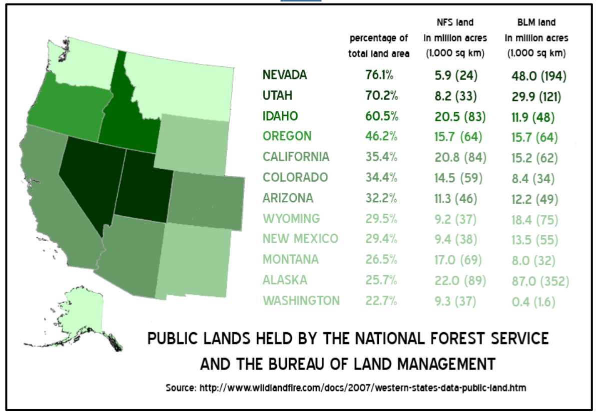
FROM: Shaun Donovan, Director
Office of Management and Budget

Christina Goldfuss, Managing Director
Council on Environmental Quality

John Holdren, Director
Office of Science and Technology Policy

SUBJECT: Incorporating Ecosystem Services into Federal Decision Making

Overview. Nature provides vital contributions to economic and social well-being that are often not traded in markets or fully considered in decisions. This memorandum provides direction to agencies on incorporating ecosystem services into Federal planning and decision making. (Broadly defined, ecosystem services are the benefits that flow from nature to people, e.g., nature's contributions to the production of food and timber; life-support processes, such as water purification and coastal protection; and life-fulfilling benefits, such as places to recreate.)

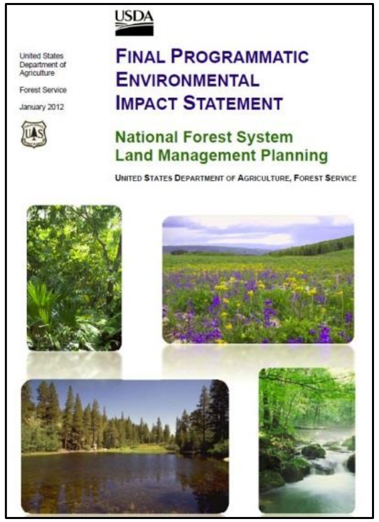


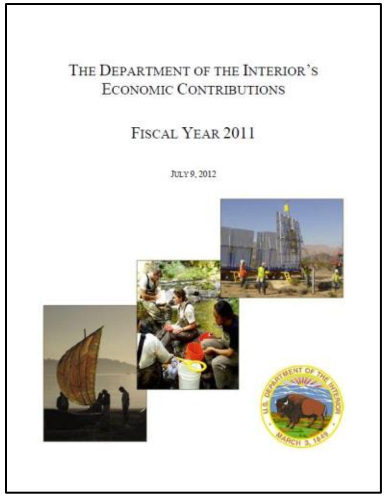
	percentage of total land area	NFS land in million acres (1,000 sq km)	BLM land in million acres (1,000 sq km)
NEVADA	76.1%	5.9 (24)	48.0 (194)
UTAH	70.2%	8.2 (33)	29.9 (121)
IDAHO	60.5%	20.5 (83)	11.9 (48)
OREGON	46.2%	15.7 (64)	15.7 (64)
CALIFORNIA	35.4%	20.8 (84)	15.2 (62)
COLORADO	34.4%	14.5 (59)	8.4 (34)
ARIZONA	32.2%	11.3 (46)	12.2 (49)
WYOMING	29.5%	9.2 (37)	18.4 (75)
NEW MEXICO	29.4%	9.4 (38)	13.5 (55)
MONTANA	26.5%	17.0 (69)	8.0 (32)
ALASKA	25.7%	22.0 (89)	87.0 (352)
WASHINGTON	22.7%	9.3 (37)	0.4 (1.6)

PUBLIC LANDS HELD BY THE NATIONAL FOREST SERVICE AND THE BUREAU OF LAND MANAGEMENT

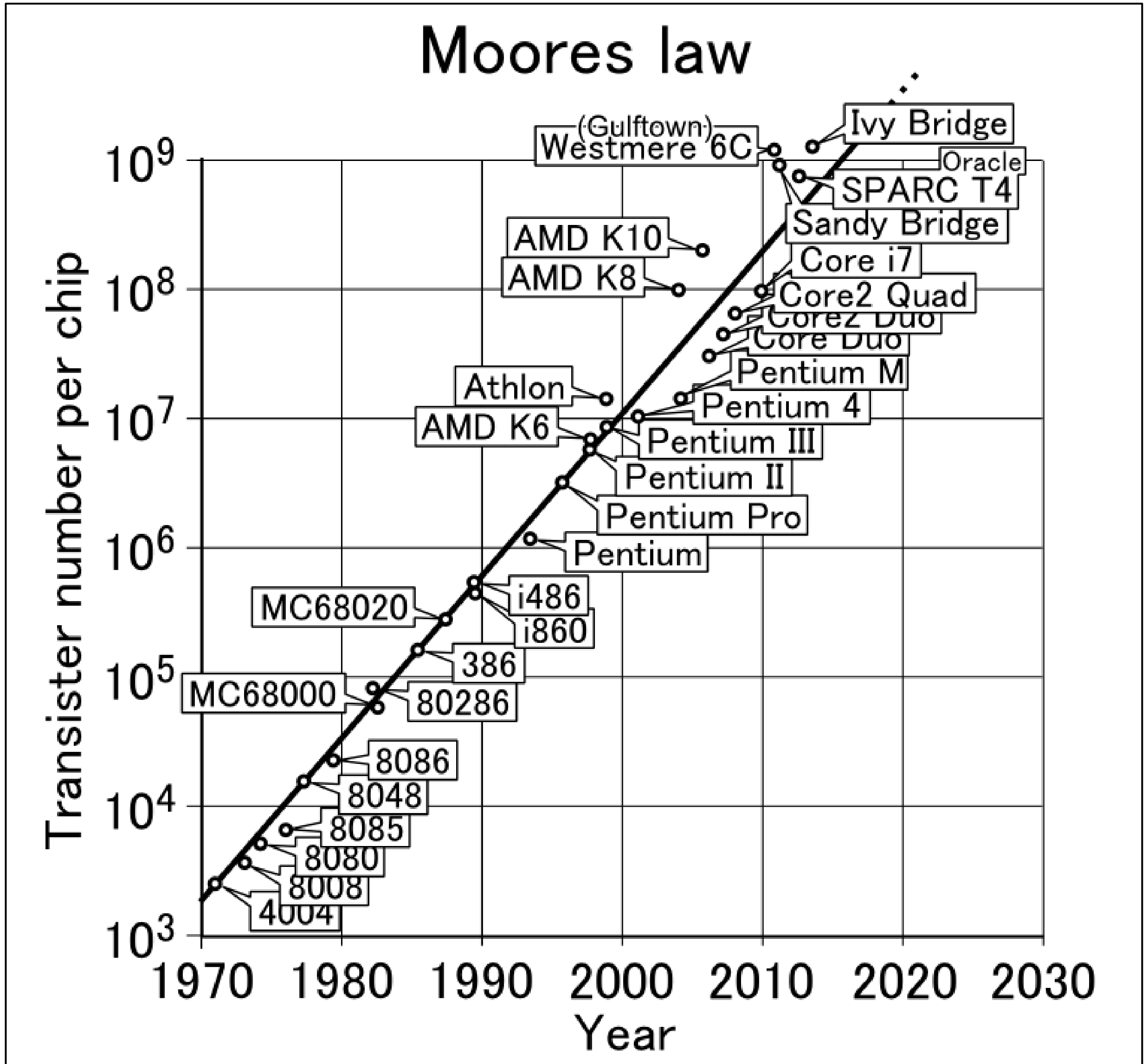
Source: http://www.wildlandfire.com/docs/2007/western-states-data-public-land.htm

Same, *but* short on resources & in-depth technical expertise





2. Rise of computing power & computer science

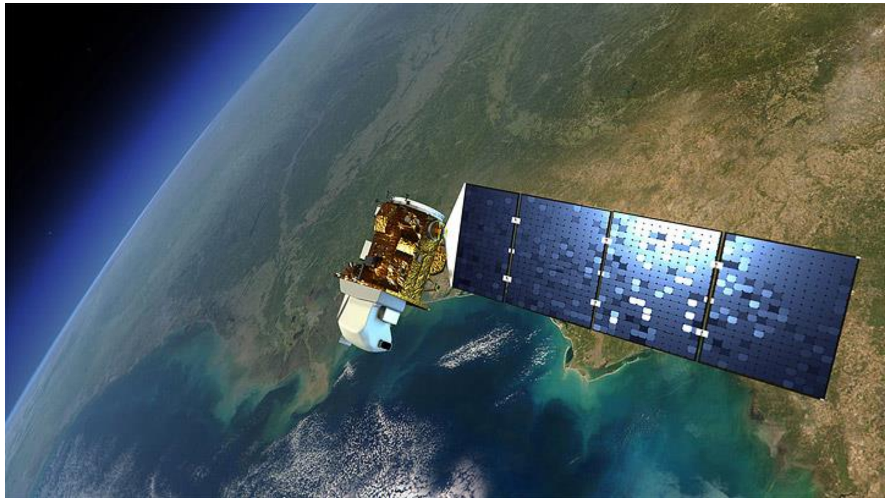


2. Rise of computing power & computer science

Google Cloud Platform

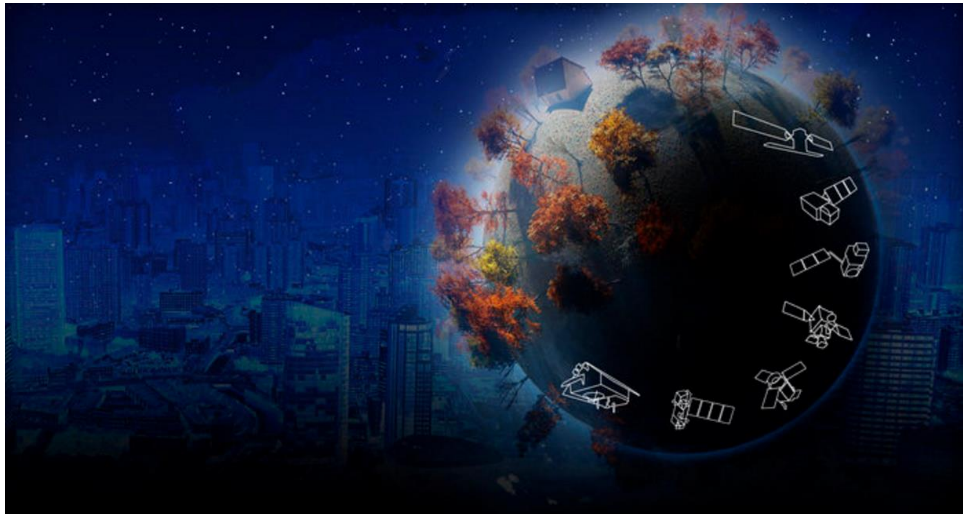
amazon web services

3. Proliferation of new data sources



Landsat 8

"There's no question we're in a golden age for remote sensing."
- Michael Freilich, NASA Earth Science Director



https://sentinel.esa.int/web/sentinel/home

http://www.nextpowerup.com/news/22115/oneweb-launching-satellite-based-internet-in-2019/

3. Proliferation of new data sources

http://rmgsc.cr.usgs.gov/uas/

3. Proliferation of new data sources

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
SOCIOECONOMIC SCENARIOS
CONSERVATION
ENVIRONMENTAL ECONOMICS
ECOSYSTEM SERVICES

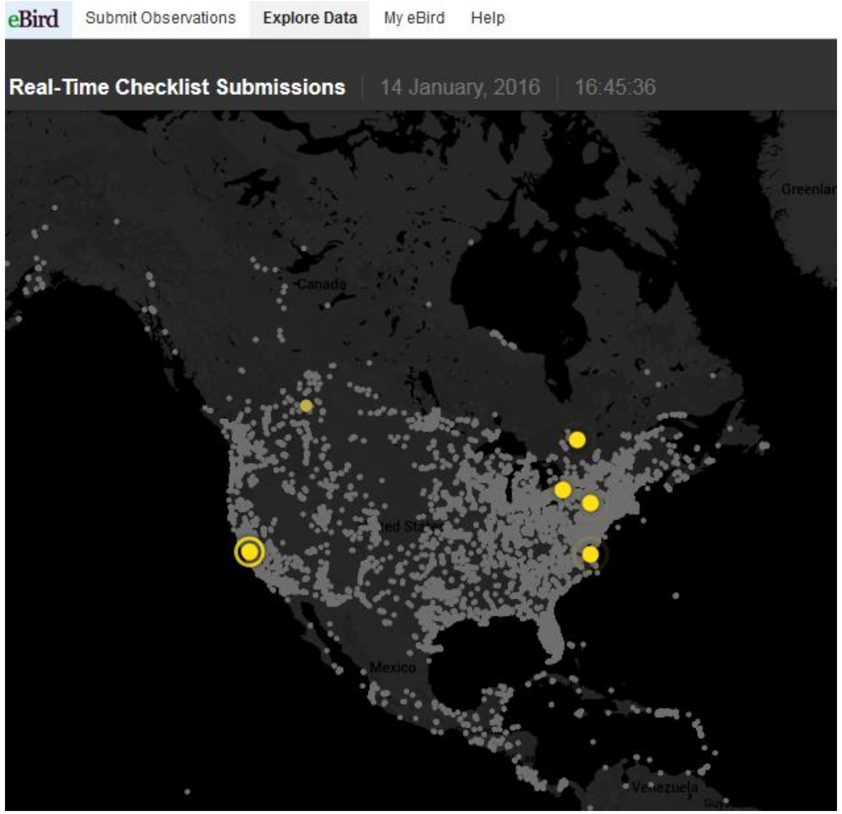
Received

Using social media to quantify nature-based tourism and recreation

Spencer A. Wood[1,2], Anne D. Guerry[1,2], Jessica M. Silver[1,2] & Martin Lacayo[2]

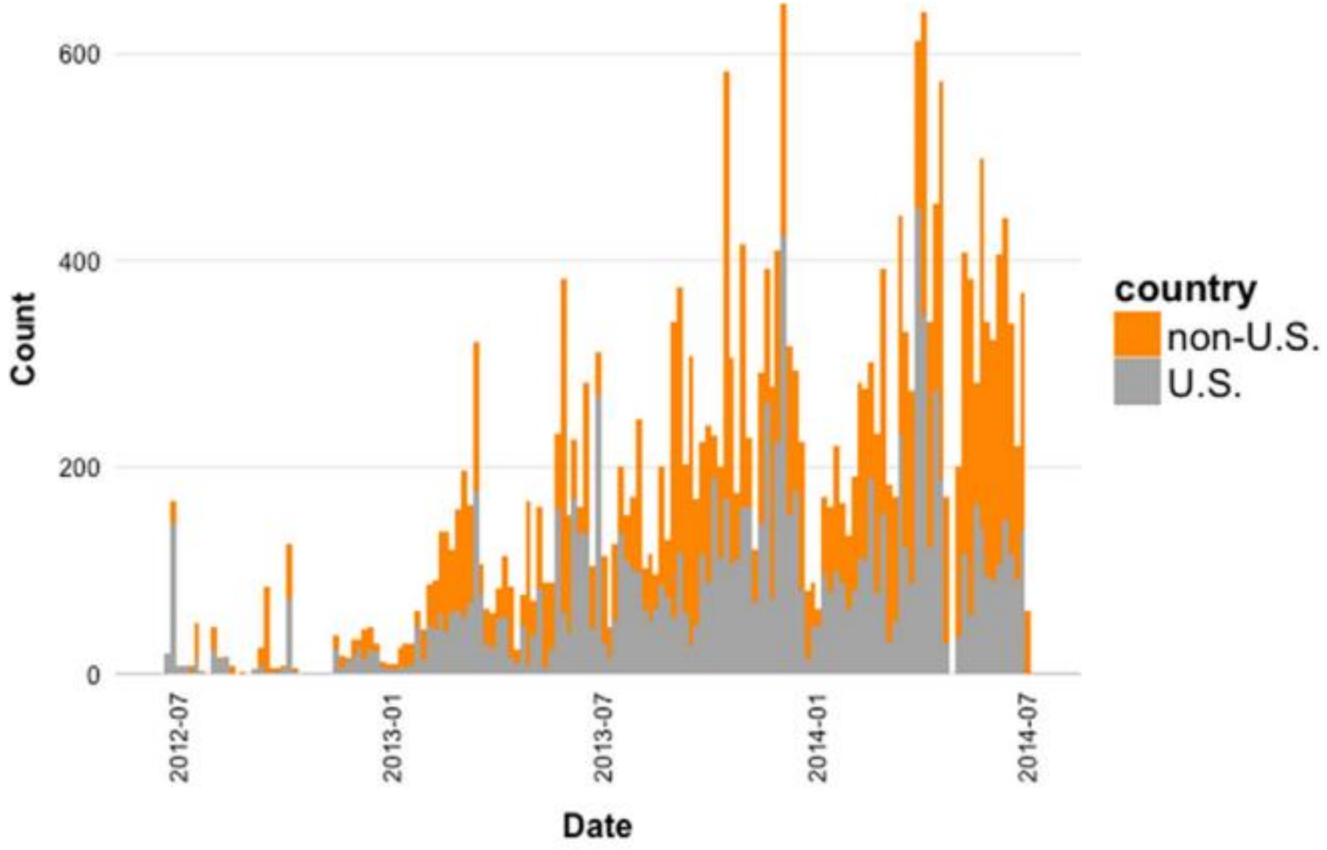
[1]Natural Capital Project, School of Environmental and Forest Sciences, University of Washington, Seattle, WA, USA, [2]Natural Capital Project, Woods Institute for the Environment, Stanford University, Stanford, CA, USA.

Scientists have traditionally studied recreation in nature by conducting surveys at entrances to major attractions such as national parks. This method is expensive and provides limited spatial and temporal



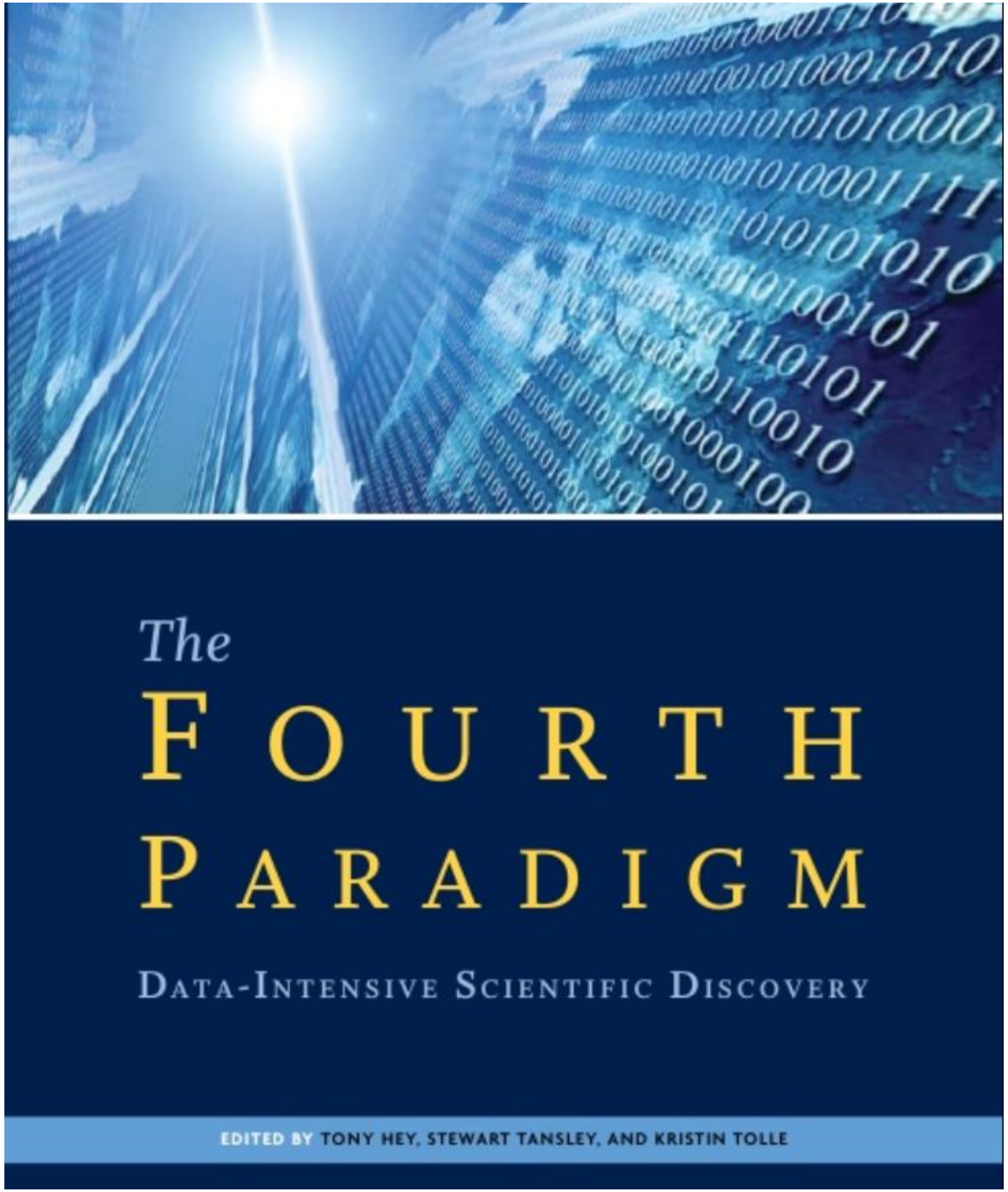
4. We're learning a tremendous amount about ecosystem services… but we're not:

- Delivering in-time information to decision makers
- Driving down the time & cost of scientifically rigorous ES analysis fast enough
- Build off existing modeling studies & supporting meta analysis efficiently



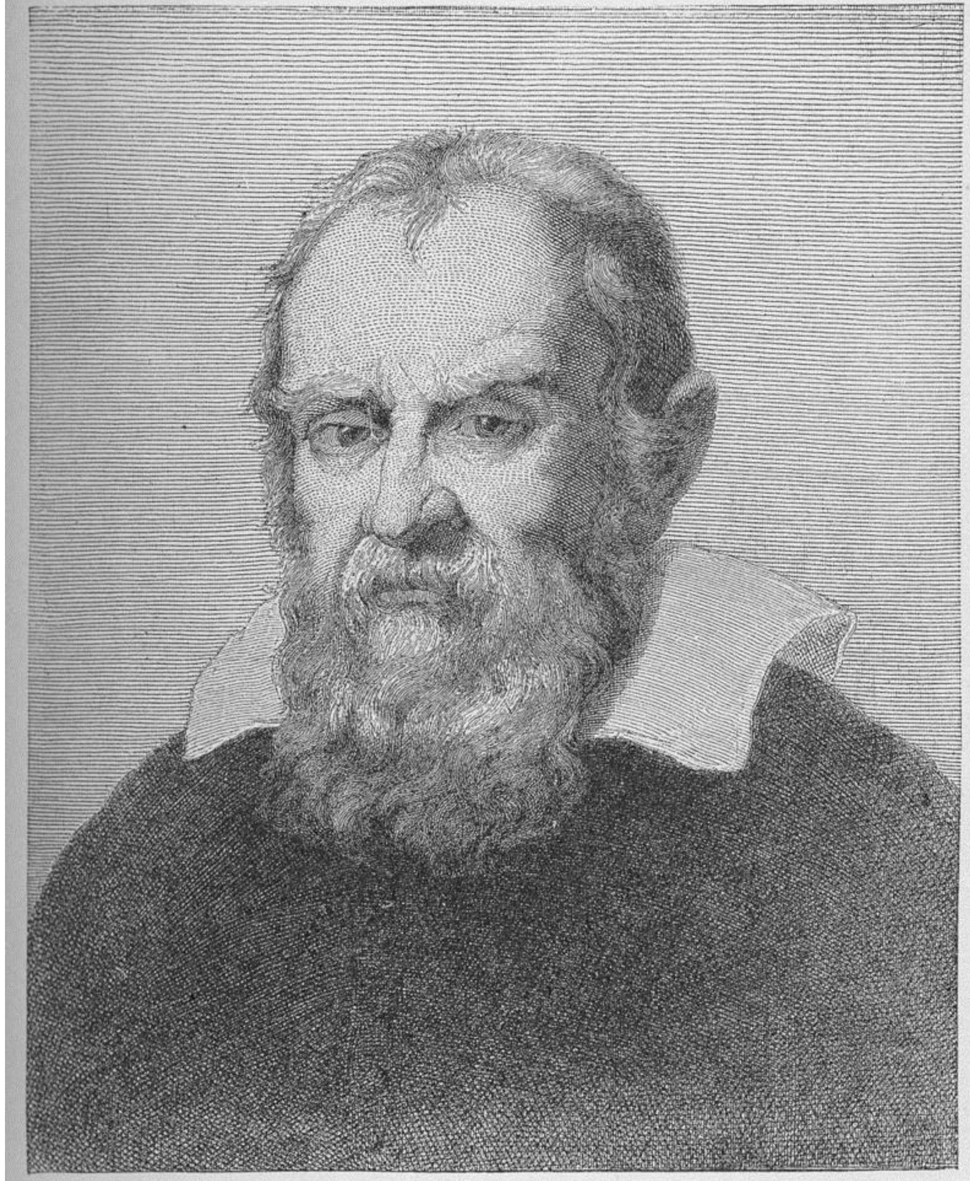
Posner et al. 2016

Data-driven
(i.e., inductive)
science
Hey et al.
2009



1. Descriptive Science

2. Theoretical Science



GALILEO GALILEI.

3. Computational Science

Deductive science

↑ Explanation why

↓ Explanatory power (r^2)

Good when you:

Have strong understanding of how a system works

Want to test why a system works the way it does

Pitfalls:

Sloppy model construction

Not following best practices in modeling

4. Data-Intensive Science

"Schmidt-Brin-Page-20080520" by Joi Ito from Inbamura, Japan - Eric Schmidt, Sergey Brin and Larry Page. Licensed under CC BY 2.0 via Wikimedia Commons - https://commons.wikimedia.org/wiki/File:Schmidt-Brin-Page-20080520.jpg#/media/File:Schmidt-Brin-Page-20080520.jpg

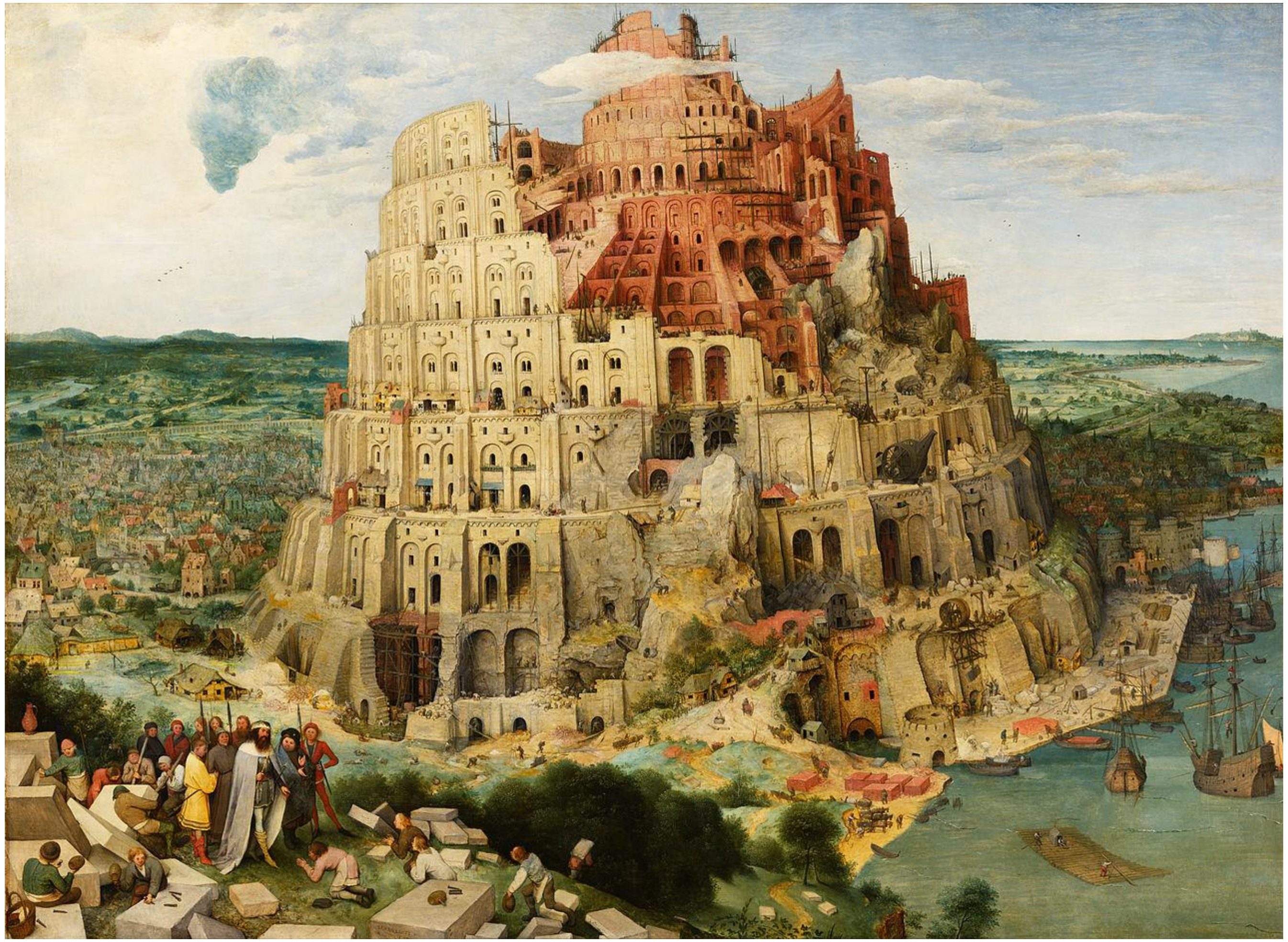
Inductive science

↑ Explanatory power (r^2)
↓ Explanation why

Good when you:
Want to explore for patterns
Want to see if real-world patterns conform to theory
Have incomplete datasets or high uncertainty

Pitfalls:
Putting too much faith into patterns found in the data that lack a reasonable theoretical foundation

Semantics in modeling



Group

Period	I	II											III	IV	V	VI	VII	VIII
1	1 H																	2 He
2	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	*	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	**	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Uut	114 Fl	115 Uup	116 Lv	117 Uus	118 Uuo
8	119 Uun																	

* Lanthanides	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu
** Actinides	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr

CAROLI LINNÆI, SVECI,

DOCTORIS MEDICINÆ.

SYSTEMA NATURÆ,

SIVE

REGNA TRIA NATURÆ

SYSTEMATICE PROPOSITA

PER

CLASSES, ORDINES, GENERA, & SPECIES.

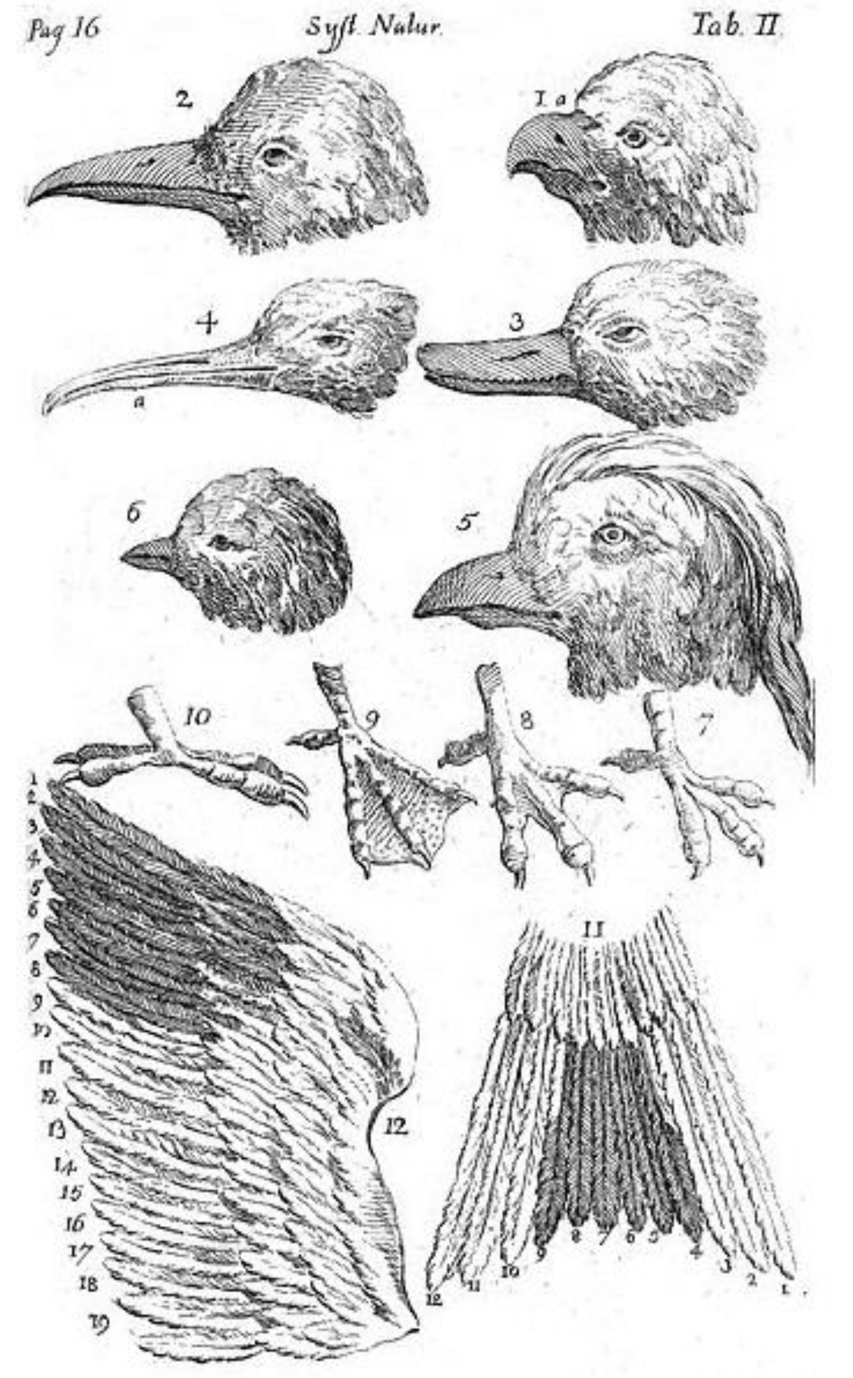
O JEHOVA! Quam ampla sunt opera Tua!
Quam ea omnia sapienter fecisti!
Quam plena est terra possessione tua!

Psalm. civ. 24.

LUGDUNI BATAVORUM,

Apud THEODORUM HAAK. MDCCXXXV.

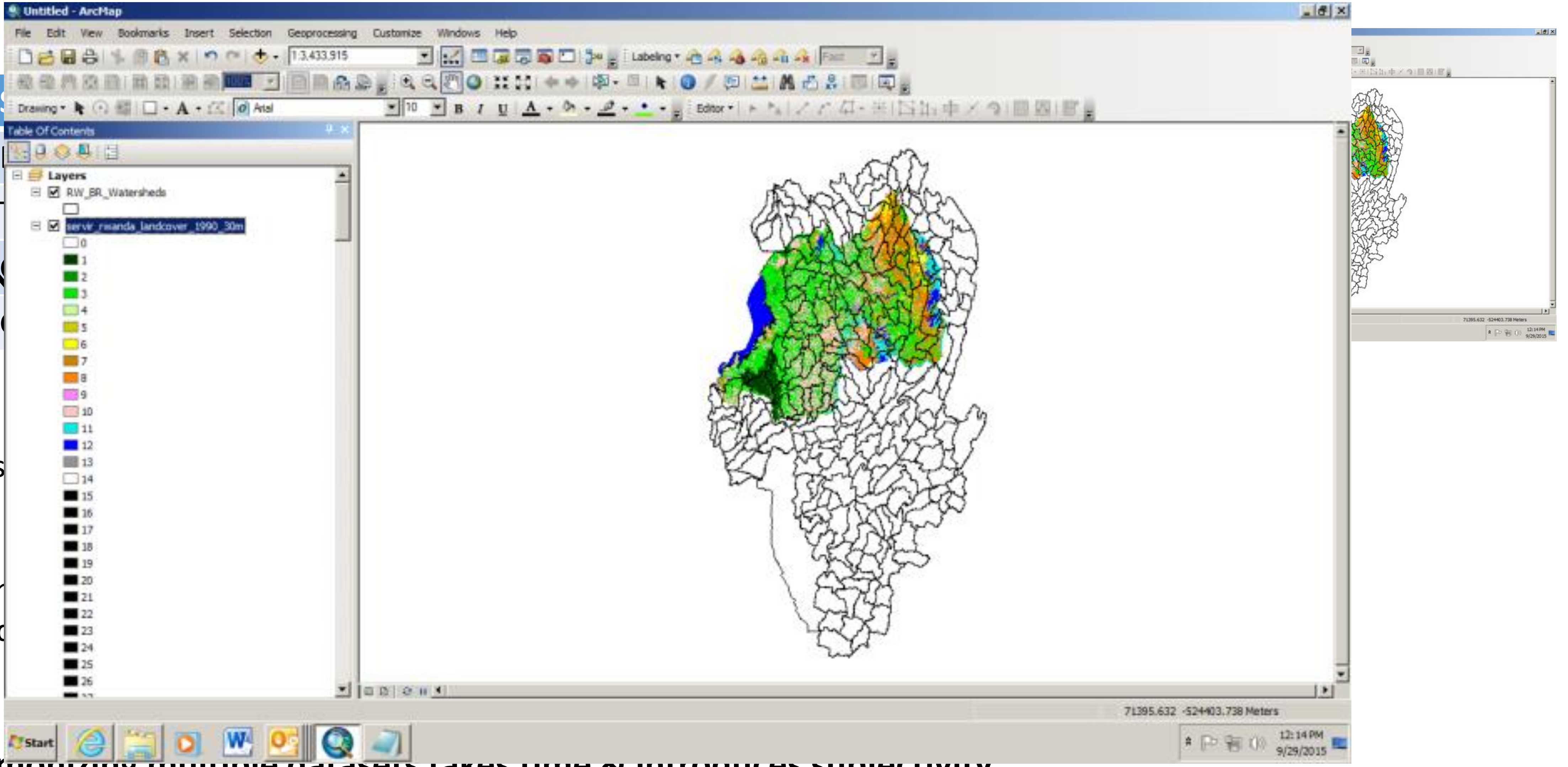
Ex Typographia
JOANNIS WILHELMI DE GROOT.



Classifying land cover: A semantic jungle

Data source
SERVIR
SERVIR
ESA-CCI
GlobeLand

Dense forest
Moderate fores
Sparse forest
Woodland
Closed shrublan
Open shrubland

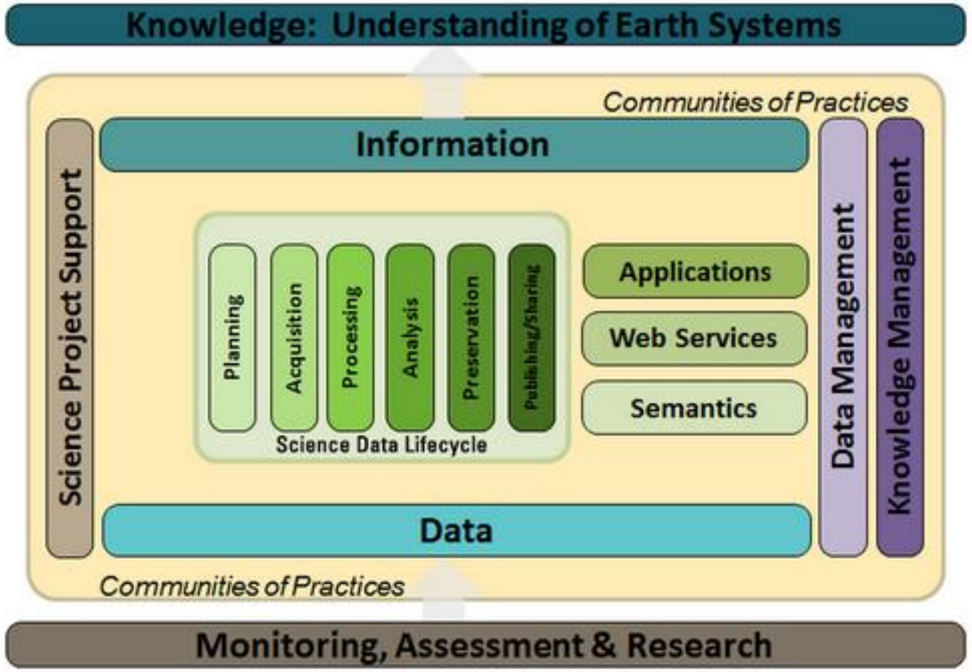


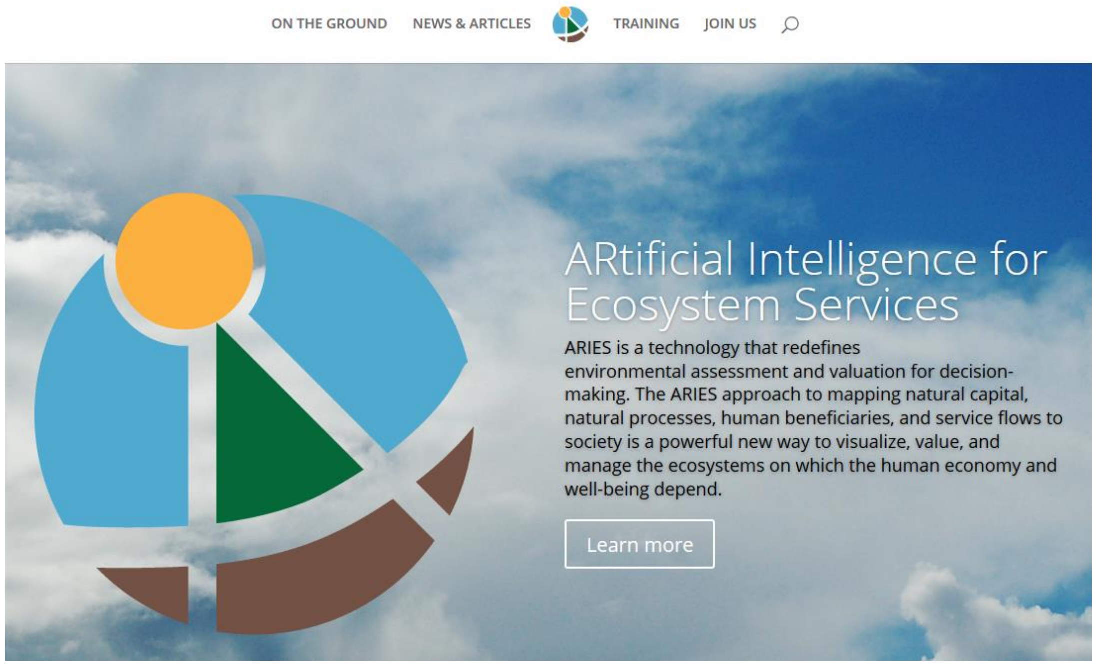
<u>Manually harmonizing multiple datasets takes time & introduces subjectivity</u>

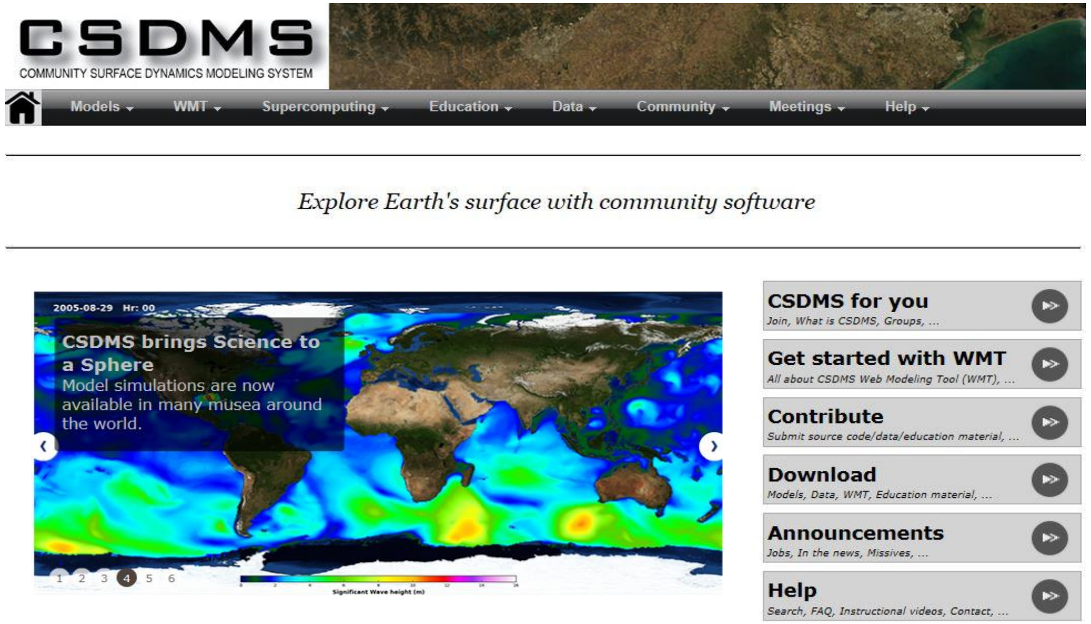
Semantics: The way out of the data & models jungle

USGS
science for a changing world

NASA
Jet Propulsion Laboratory
California Institute of Technology
SWEET
Semantic Web for Earth and Environmental Terminology







A semantic approach to ecosystem services (Villa et al. 2014)

Key design principles for semantic modeling

1. Keep semantics as modular as possible

2. Reuse existing expert knowledge wherever possible

3. Formalize semantics in *ontologies**

*Concepts & their interrelationships precisely defined

Anything we can observe (with data) has a *subject*

- Countable, physical, recognizable object
- Examples:
    1. A mountain
    2. A population of humans
    3. A population of trees (i.e., a forest)
    4. A river




Typical data describe a subject's specific *quality*

- Described by an *observer type* (e.g., measurement, count, percentage, proportion…)
- Examples:
    1. The elevation of a mountain (measurement)
    2. Per capita income of a group of humans (value)
    3. Percent tree canopy cover (percent)
    4. A river's stream order (ranking)




Over time, subjects may experience ***processes***

- Examples:
  1. Erosion of a mountainside
  2. Migration of a human population
  3. Tree growth in a forest
  4. Streamflow in a river




A single, time-limited process is an *event*

- Examples:
    1. A snowfall event on a mountain
    2. The birth of a new human in the population
    3. The death of a tree in the forest
    4. A flood event on a river




Relationships connect two subjects

- Examples:
  - Structural connection – Parenthood connects parents to children
  - Functional connection – Ecosystems providing a particular benefit to human beneficiaries
- Very important for agent-based models




Observables can also have one or more *traits*

- "Adjectives" that add descriptive power to further modify a concept
- Add flexibility without adding more complexity to the ontologies
- Four types:

1. Attributes (Temporal, frequency, min/max/mean, etc.)
2. Identities (Authoritative species or chemical names)
3. Strata (Soil, atmosphere, ocean, forest)
4. Orderings (High-Moderate-Low)

A language to define data for environmental modeling

Subjects

Qualities

Processes

Events

Relationships

Traits

Why hasn’t data-intensive, cloud-based ecosystem service modeling taken off on its own?

Why is cloud-based modeling hard?

Why is cloud-based modeling hard?

Advertising

Health care

Environment

Private goods

Public goods

Why is cloud-based modeling hard?

I don’t understand the value of big data!

My data are more valuable to me if I keep others from using them!

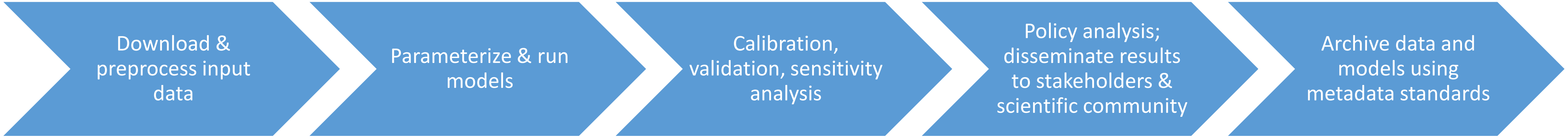
Skeptic 1 Data scientist Skeptic 2

Why is cloud-based modeling hard?

Some proposals on where to start

Standard data & modeling flow

- First time running an environmental modelers:



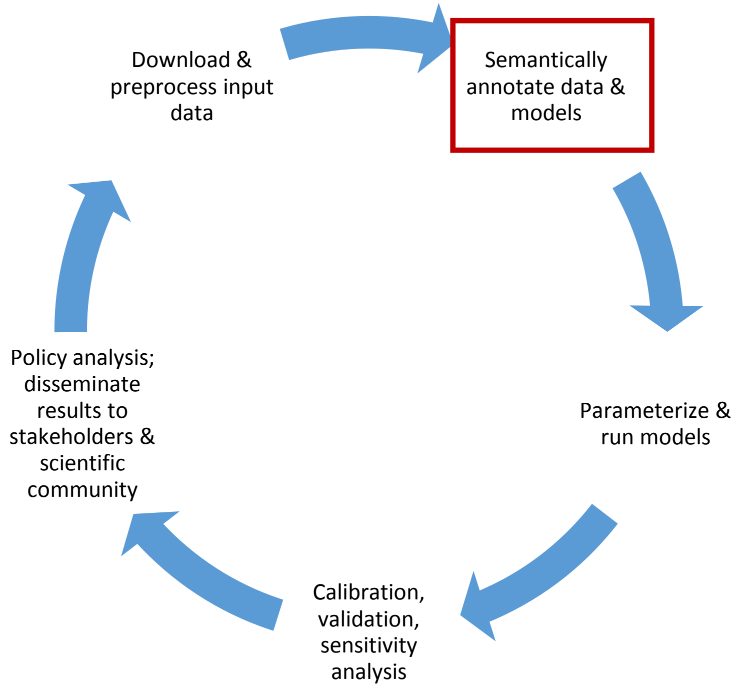
- Second, third, fourth time:

Same thing!

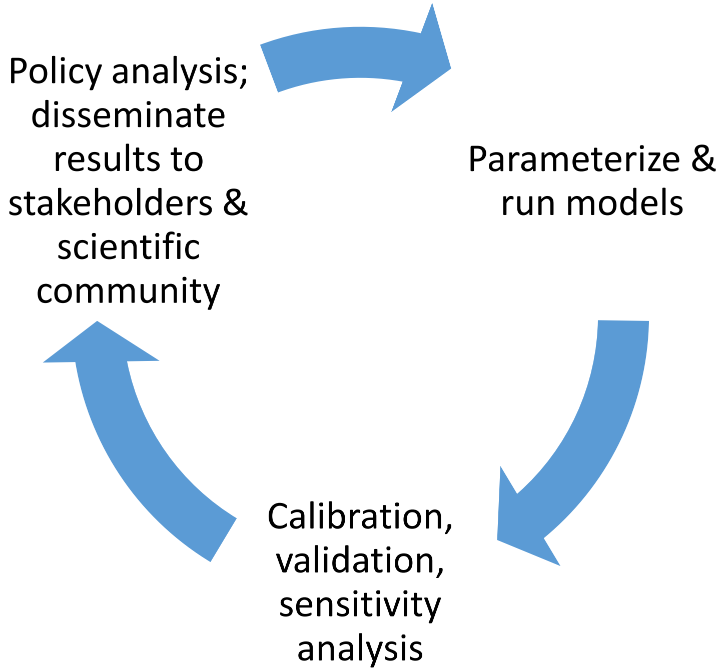
At the end of the project, budgets are tight, and people want to get the paper/report out without worrying about proper archiving

Semantic data & modeling flow

- First time modelers:



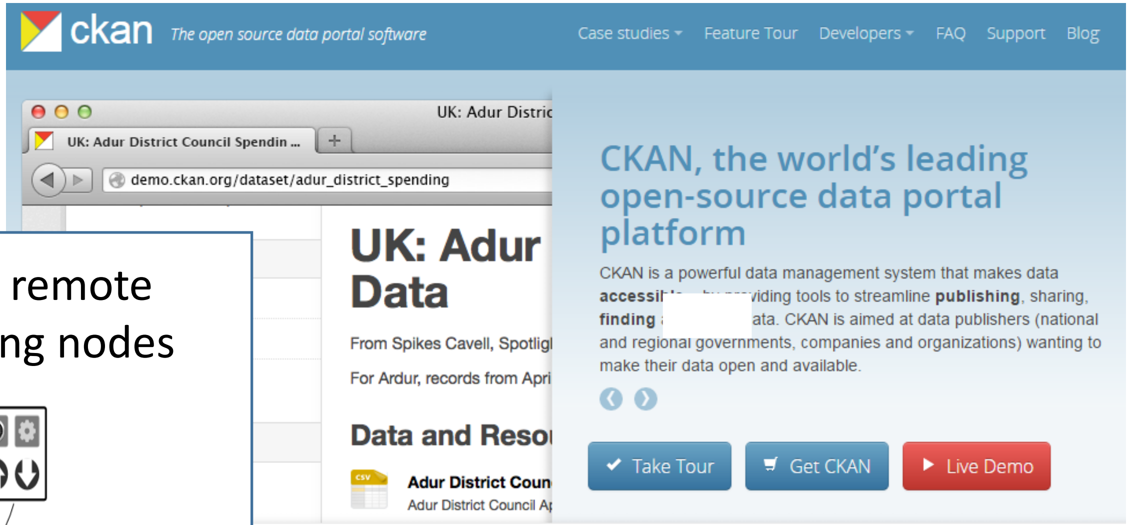
Second, third, fourth time modelers:



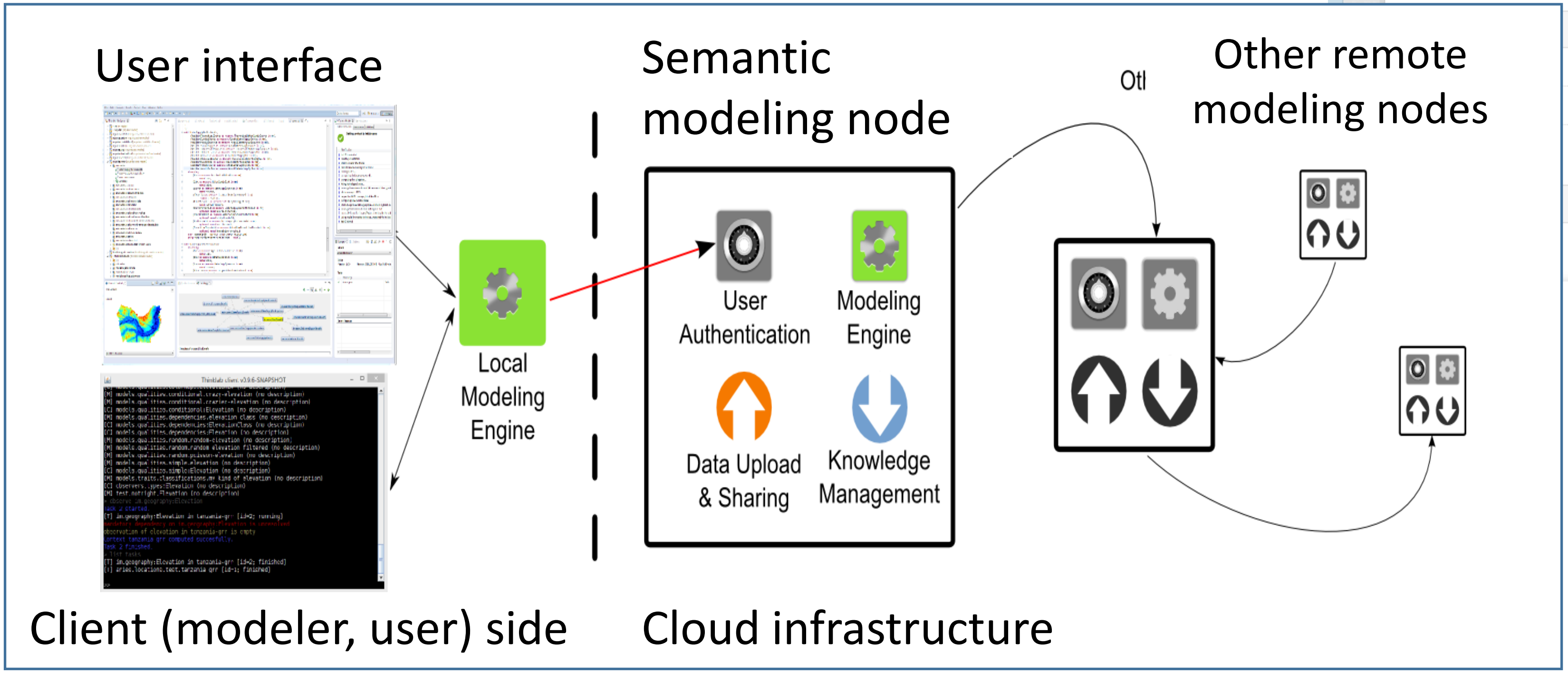
- Archival work is ***up front***, after that it's usable by anyone

Cloud-based data and models: Toward <u>context-aware modeling</u>

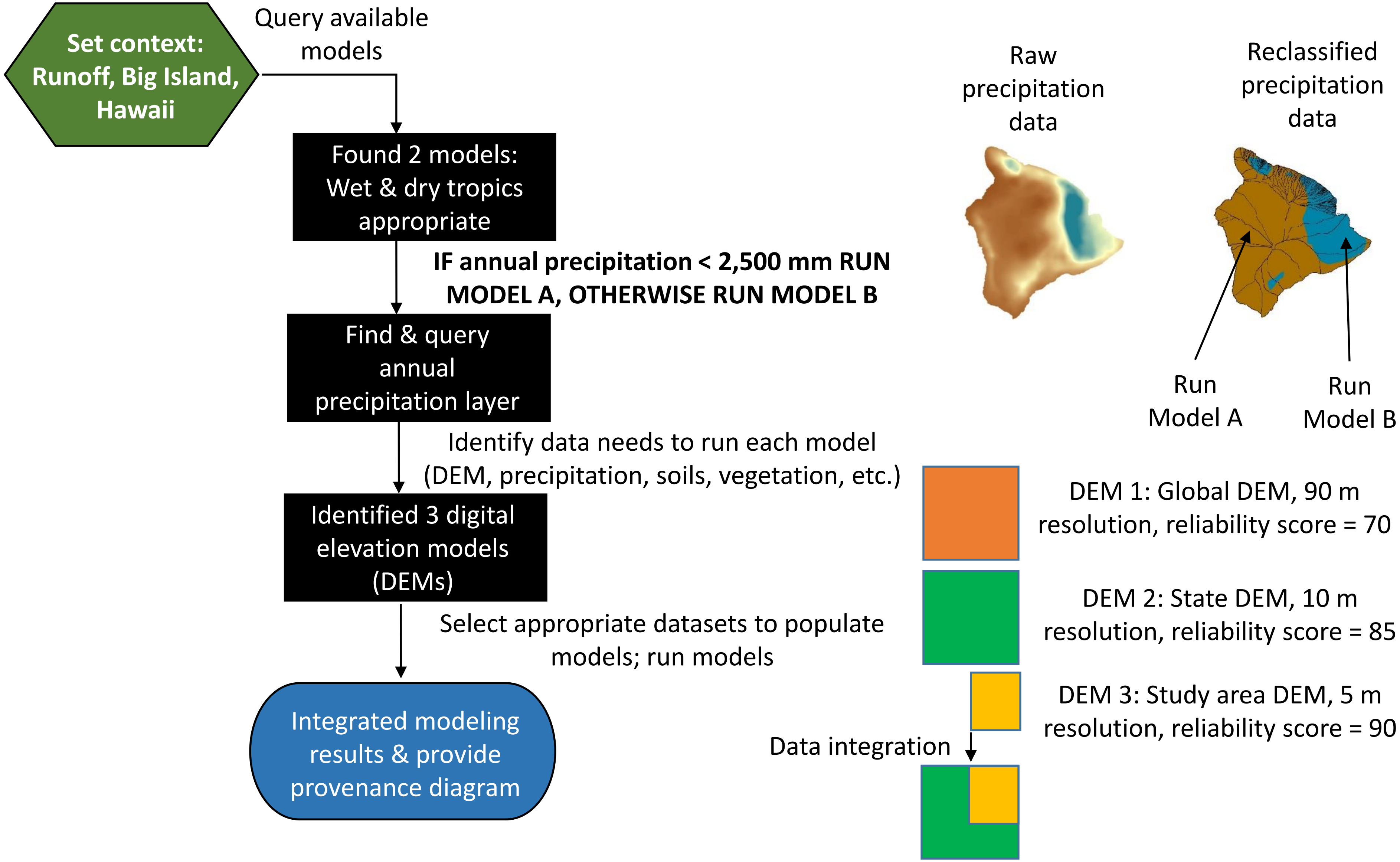
- Many global & national datasets served by Web Coverage Service (WCS)/Web Feature Service (WFS)
- Data export via CKAN



GeoServer

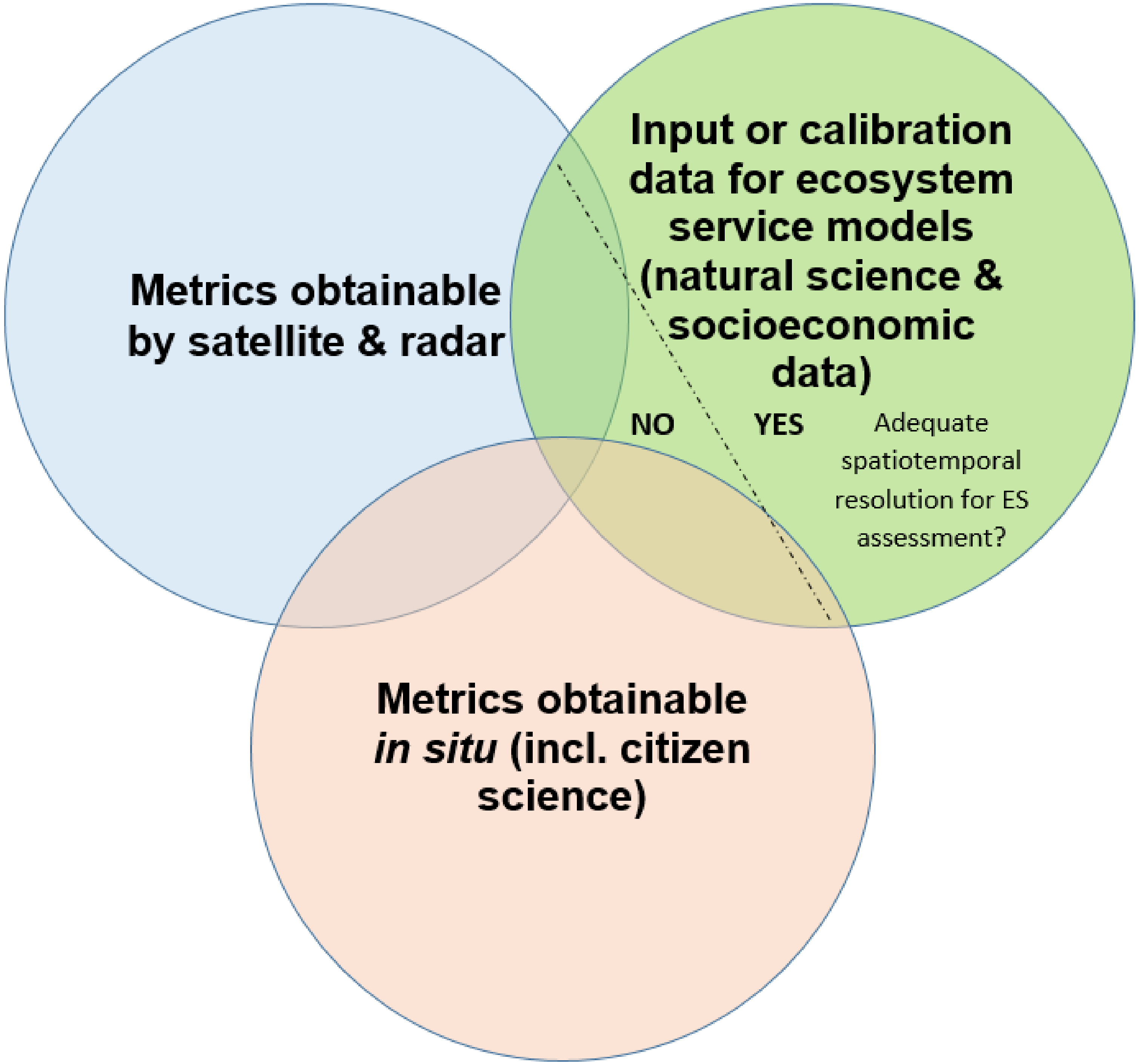


Automating data & model assembly: Opens door for machine reasoning, pattern recognition



Making smarter choices about data needs for natural capital accounting

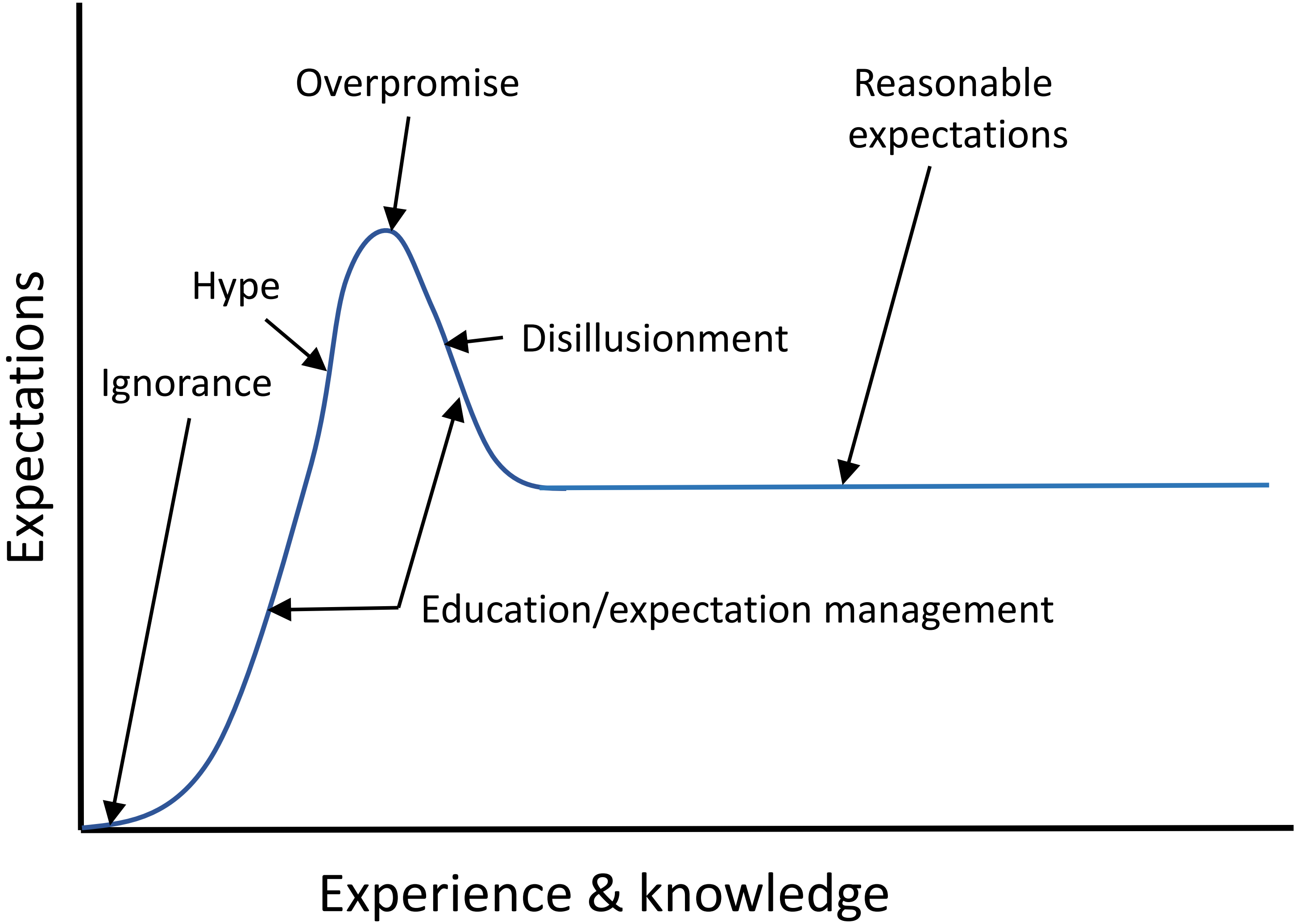
Consider needed spatiotemporal resolution and data quality



At a lunchtime plenary session at ACES 2012, Fort Lauderdale

“Who else is working on big data and semantic modeling in ecosystem services?”

Big data: Moving from hype to reality



Getting from here to there

Individual scientists & practitioners can:
1. Understand what big data is and isn't
2. Get comfortable with use of data-intensive science/inductive modeling when it's appropriate
3. Collaborate on and use semantics
4. Share, serve, annotate data to facilitate cloud-based semantic modeling
5. Fund computer science-based working on ecosystem services

Attributes of a cloud-based, big data ecosystem service modeling system
1. Be modular & independently expandible
2. Be semantically consistent
3. Support multiple modeling paradigms
4. Support context-aware modeling (AI & machine learning), i.e., be data-driven
5. Balance speed and accuracy (quick assessments/novice users), and support for high-level modeling (scientists/advanced users)

Competing 10-year visions

	Status quo	Semantic modeling*
Time for natural capital assessment	6-12 months	1-2 months
Cost of natural capital assessment	$100K-1M	$20-50K

*Requires $10M investment in semantic standards & web services

- If ecosystem services become a national & international standard requested in everyday government decision making, can we afford the status quo?
- If not, how do we finance the needed investment in semantic modeling?

Thanks!

kbagstad@usgs.gov