ES modelling & the Semantic Web: Toward *interoperability* of ES data & models

bc³
BASQUE CENTRE
FOR CLIMATE CHANGE
Klima Aldaketa Ikergai

Interoperability

“the ability of data or tools from non-cooperating resources to integrate or work together with minimal effort”

– Wilkinson et al. 2016

How well do you think we do this today?

FAIR data principles (Wilkinson et al. 2016)

- Findable
- Accessible
- Interoperable
- Reusable

Good data repositories do much of F & A

ARIES aims to help us do better especially at I & R

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

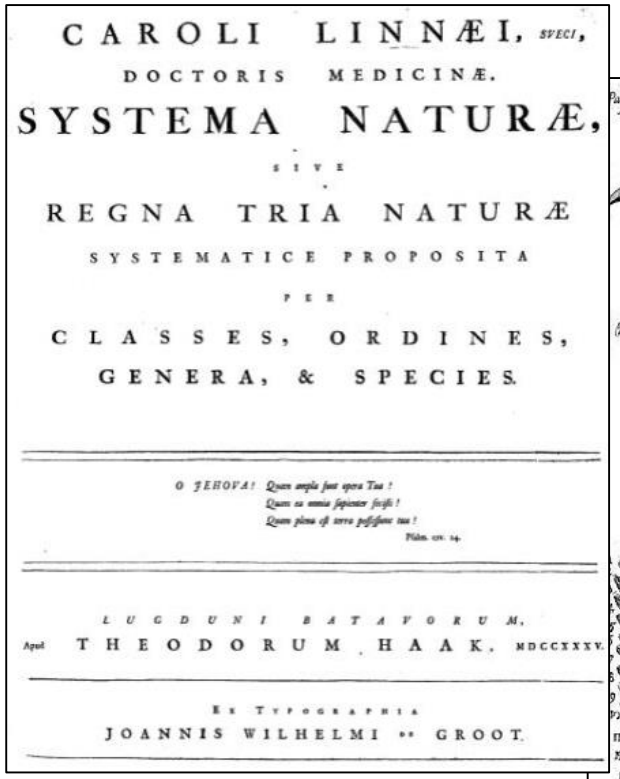
To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Is it “just semantics”?
How do we connect knowledge?

“It’s just semantics”
– Easy way to quickly end a minor disagreement about what to call something

“Data collection and preparation takes up 60% of the time needed for environmental modeling”
– Apocryphal

Classic semantics in science, with: (1) standard names, (2) relationships between things defined



CAROLI LINNÆI, SVECI,

DOCTORIS MEDICINÆ.

SYSTEMA NATURÆ,

SIVE

REGNA TRIA NATURÆ

SYSTEMATICE PROPOSITA

PER

CLASSES, ORDINES,

GENERA, & SPECIES.

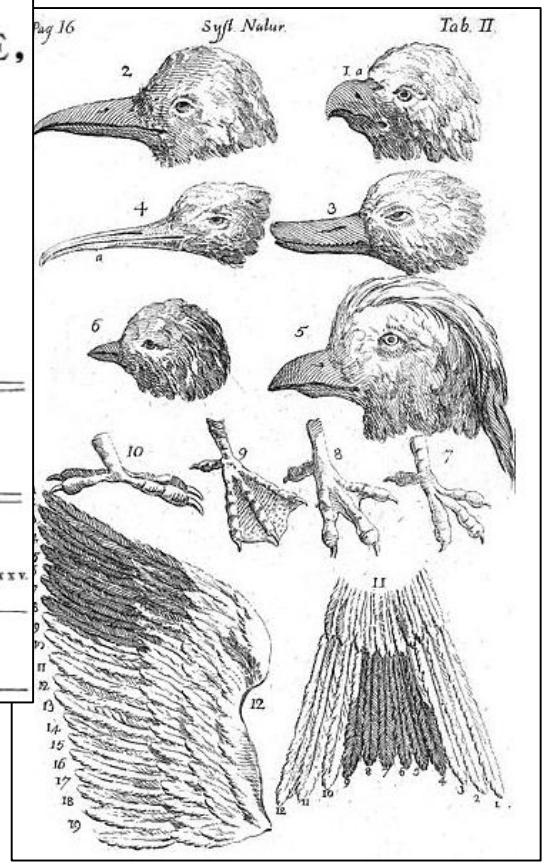
O JEHOVA! Quam ampla sunt opera Tua!
Quam ea omnia sapienter fecisti!
Quam plena est terra possessione tua!
Psalm. civ. 24.

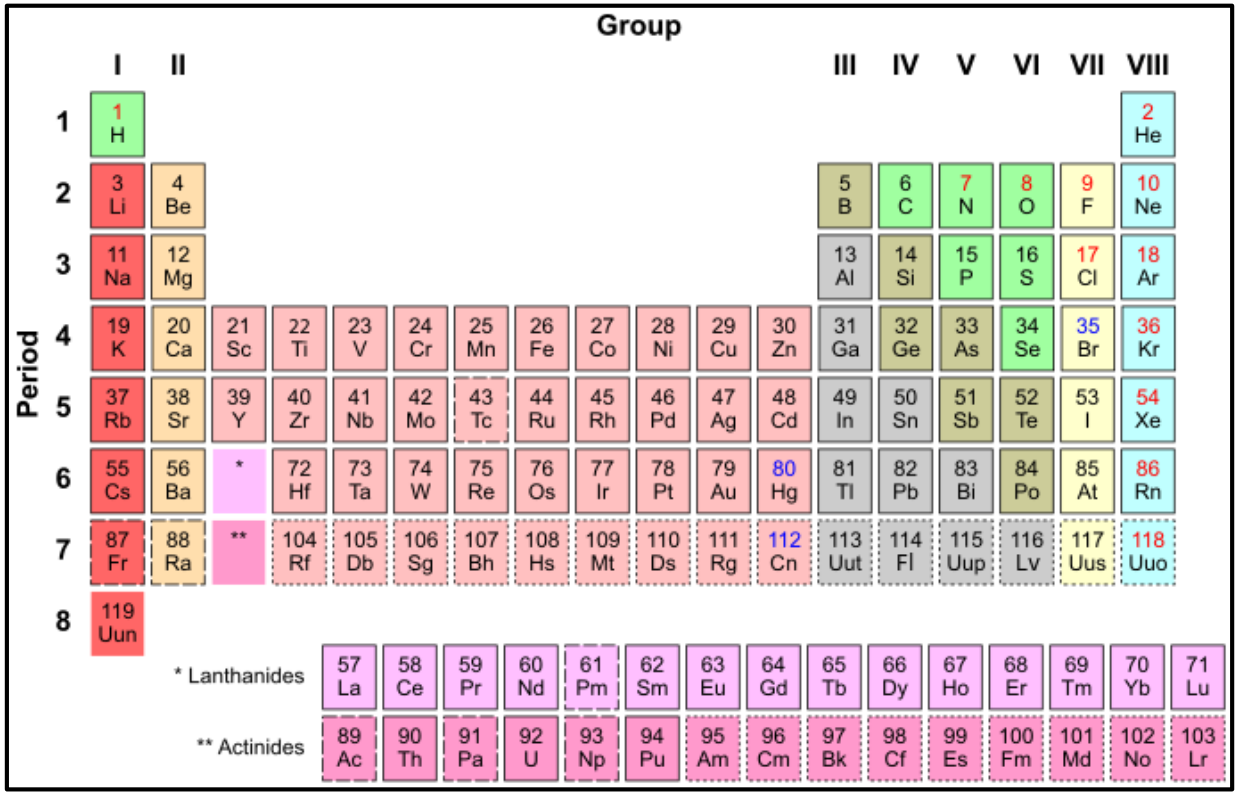
LUGDUNI BATAVORUM,

Apud THEODORUM HAAK, MDCCXXXV.

Ex Typographia

JOANNIS WILHELMI de GROOT.



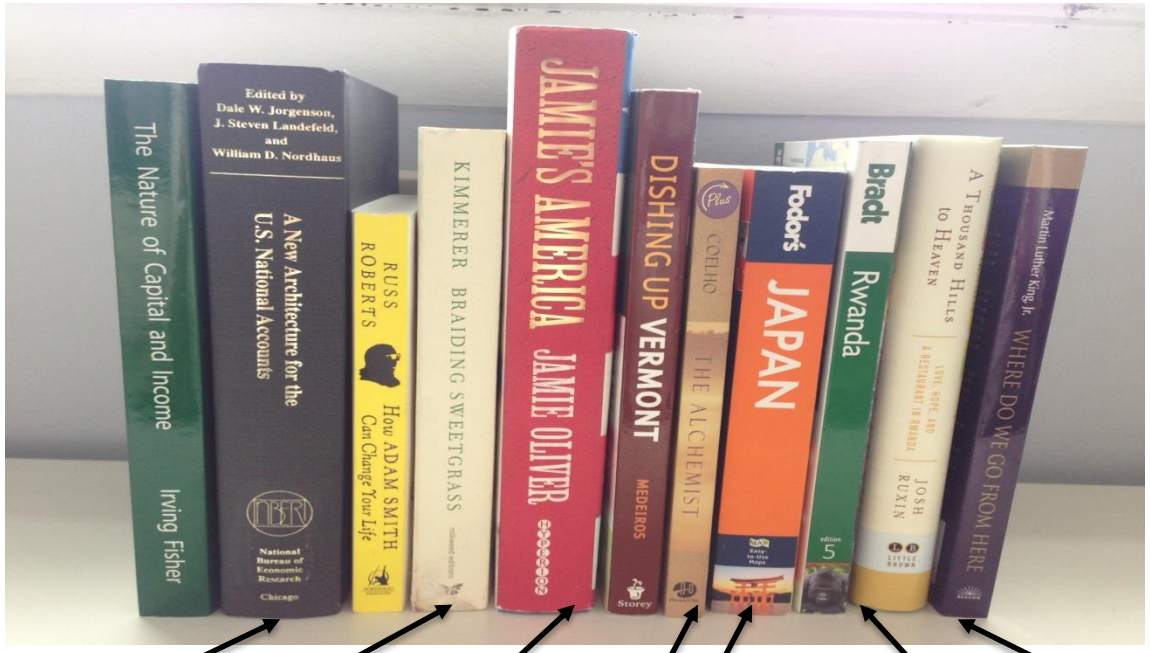


Much harder, but not impossible, to do in interdisciplinary science

Managing data semantically

How we treat (most) scientific data & models today

Advantages of treating it like this?



330 Economics

508 Natural history

641 Cooking

860 Spanish/ Portuguese Literature

915 Travel/Asia

916 Africa

973 U.S. history

How well do we share data & manage it semantically right now?

- ESP Visualization Tool (Drakou et al. 2015) has been online since January 2015, and currently hosts 30 ecosystem service maps derived from 10 studies
- Zulian et al. (2018) did a nice study on how to customize ESTIMAP models at multiple sites around the world
- InVEST keeps a database for sediment and nutrient regulation model parameters as an Excel spreadsheet summarizing past use of these models
- Great efforts, but minimal efforts to streamline/systematize data reuse
  - What are our incentives for data sharing & reuse?
  - None of these are machine readable, and we have good accepted standards for that!

www.esp-mapping.net

How well do we use semantics in science now?

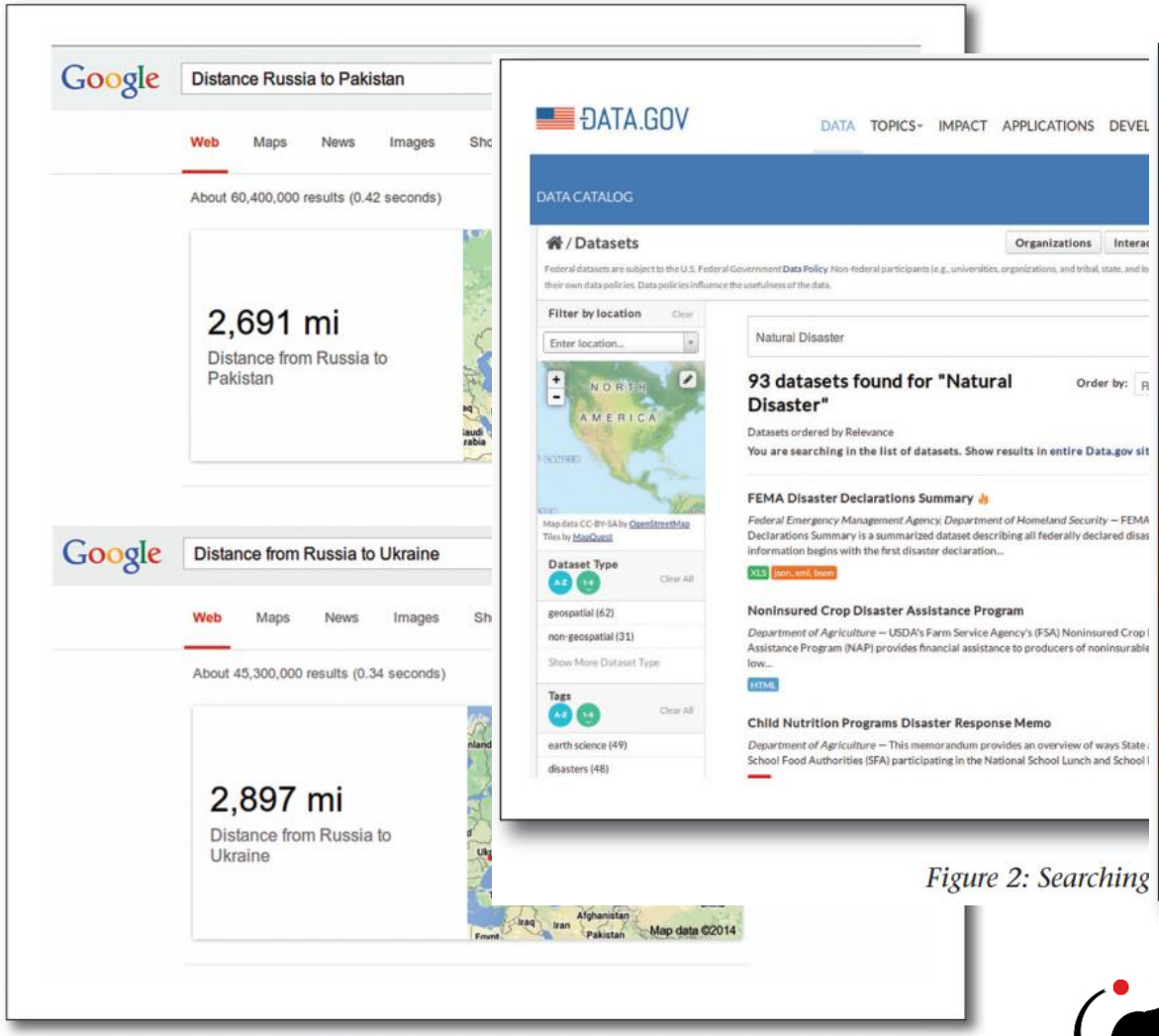


Figure 1: Distances Between Countries According to Google Search.

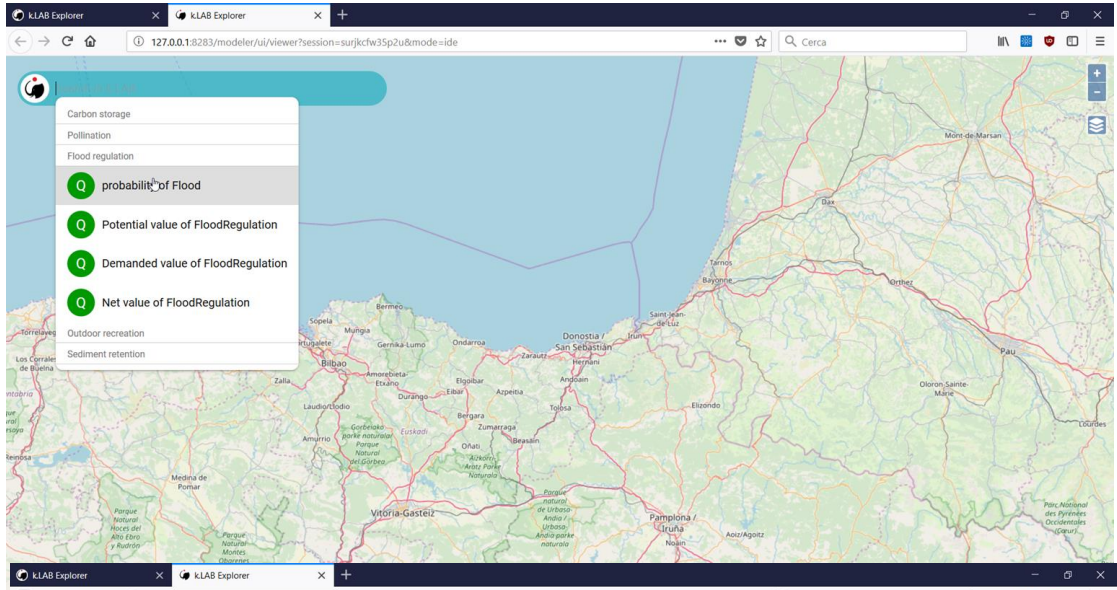
Figure 2: Searching

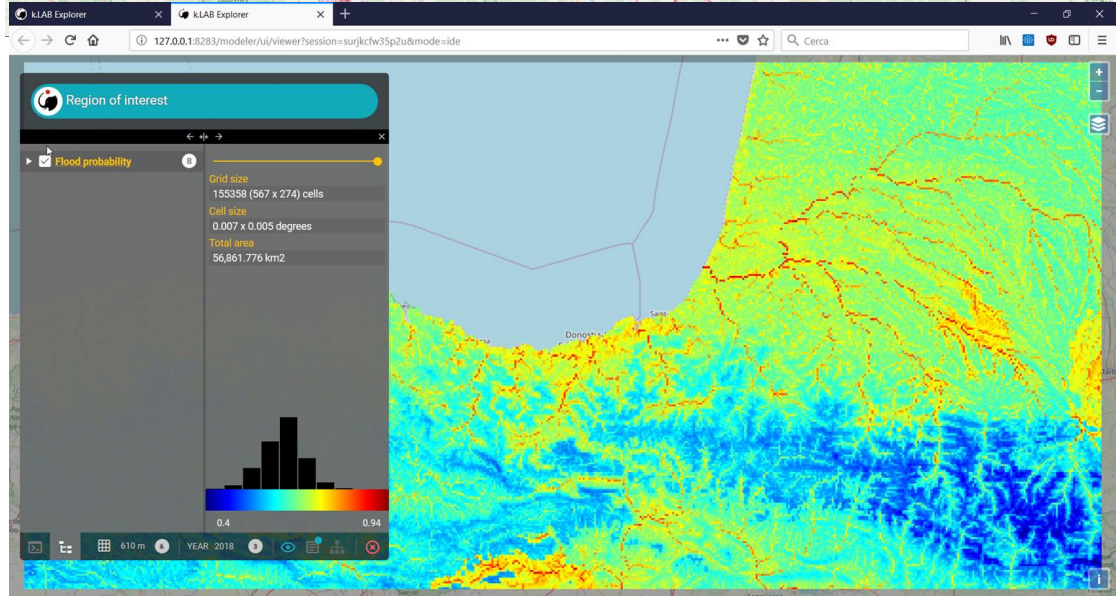
Figure 3. A Sign in New Cayama, Santa Barbara County, California, USA.

Janowicz et al. 2015 “Why the data train needs semantic rails”

Semantic web: Making data & models interoperable

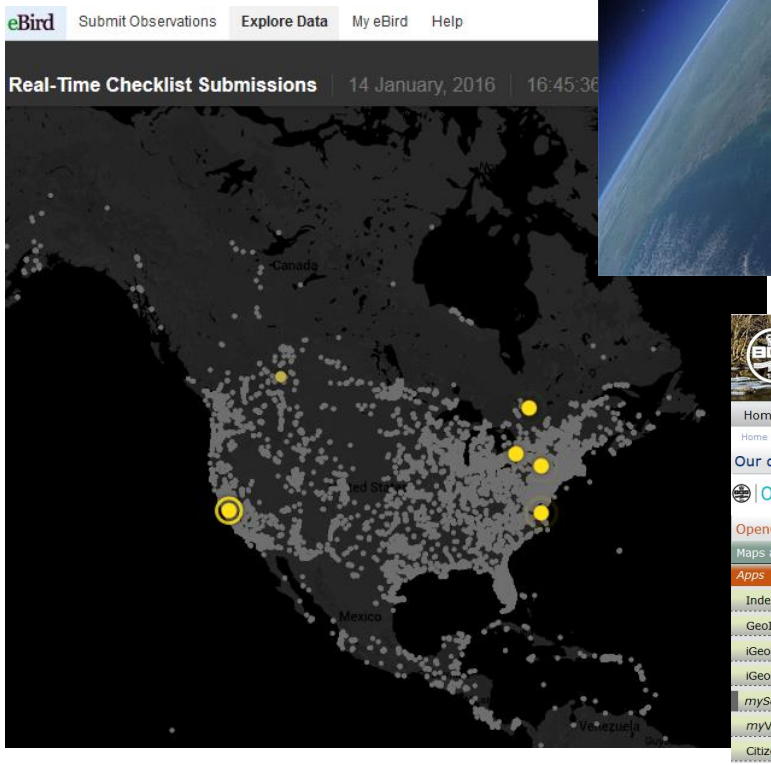
1. Put your data & models online, where they can be found by people and computers (through search)
2. Label them consistently, so people & computers know what they are, without having to make guesses
3. Develop & use apps that can assemble data & models to fit a user's needs for different places, times, and scales

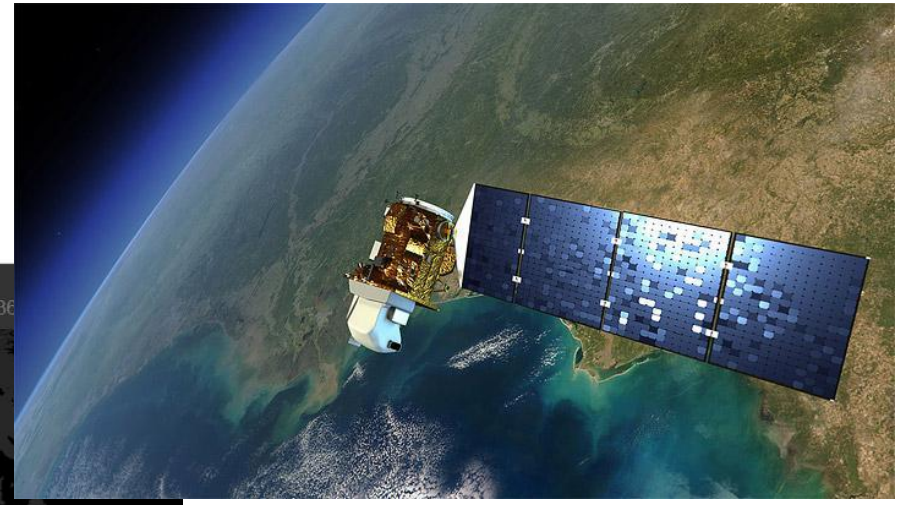




Big data

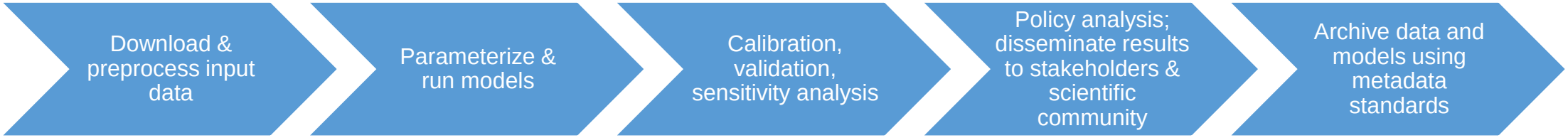
- Not just large datasets:
  - High Volume (huge amounts)
  - High Variety
  - High Velocity (coming in fast)
- How to manage the data tsunami?





Standard data & modeling flow

- First time running an environmental modelers:

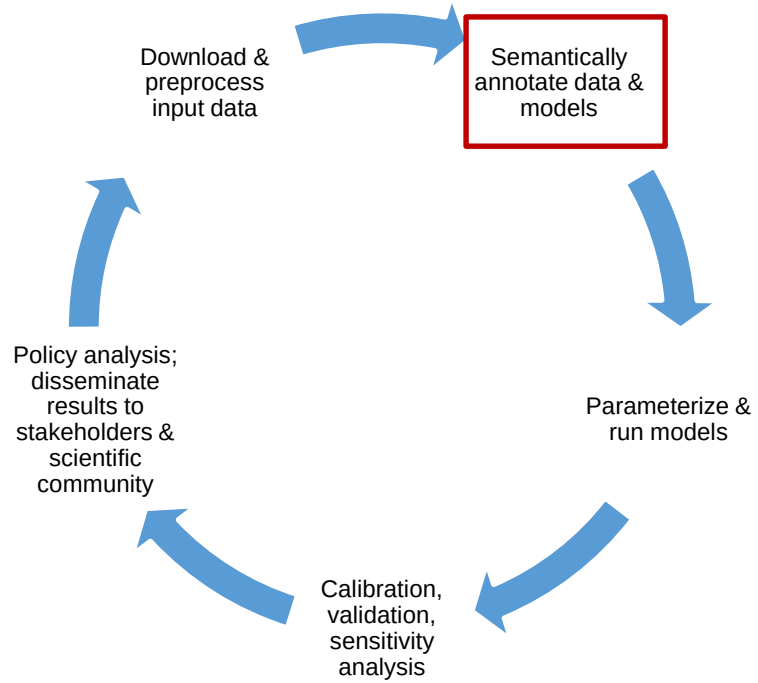


- Second, third, fourth time:

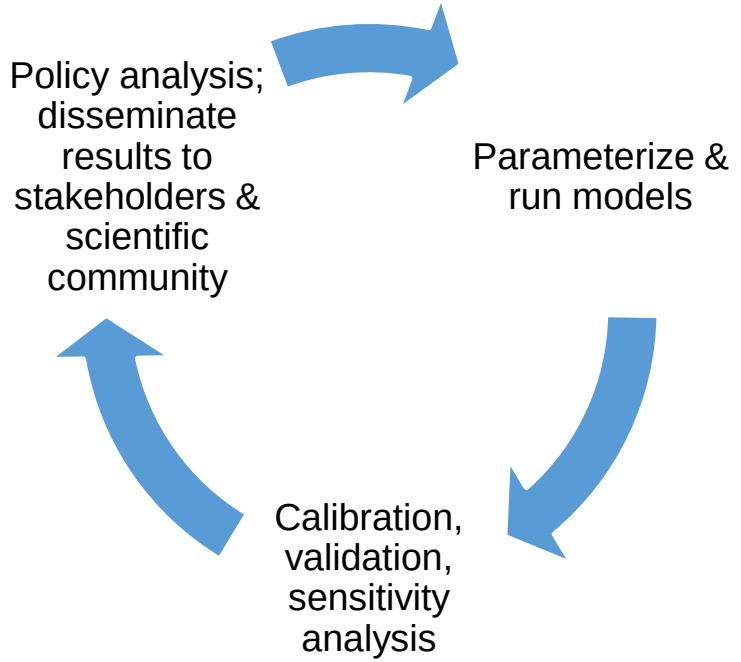
At the end of the project, budgets are tight, and people want to get the paper/report out without worrying about proper archiving

Semantic data & modeling flow

- First time modelers:



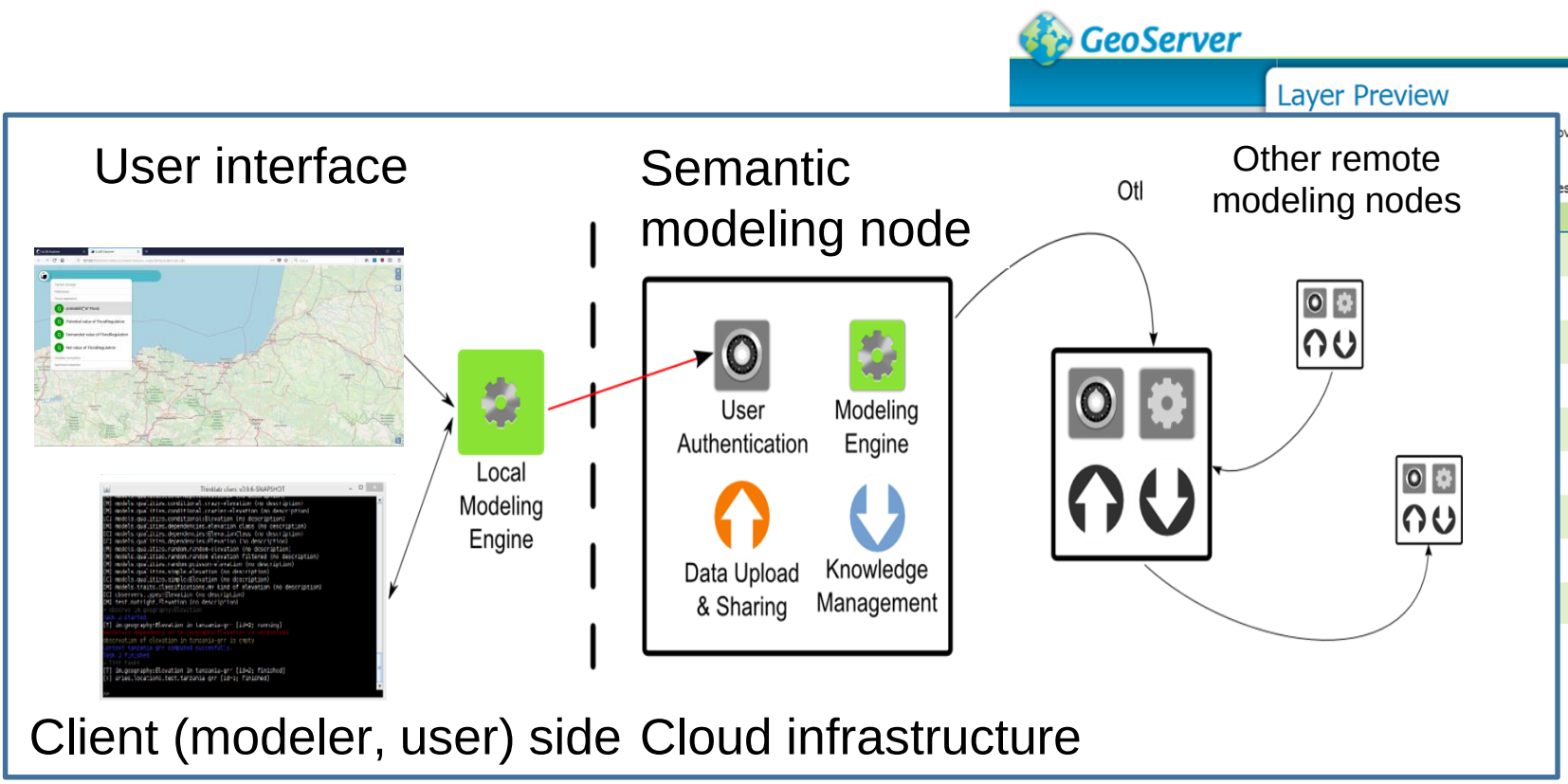
Second, third, fourth time modelers:

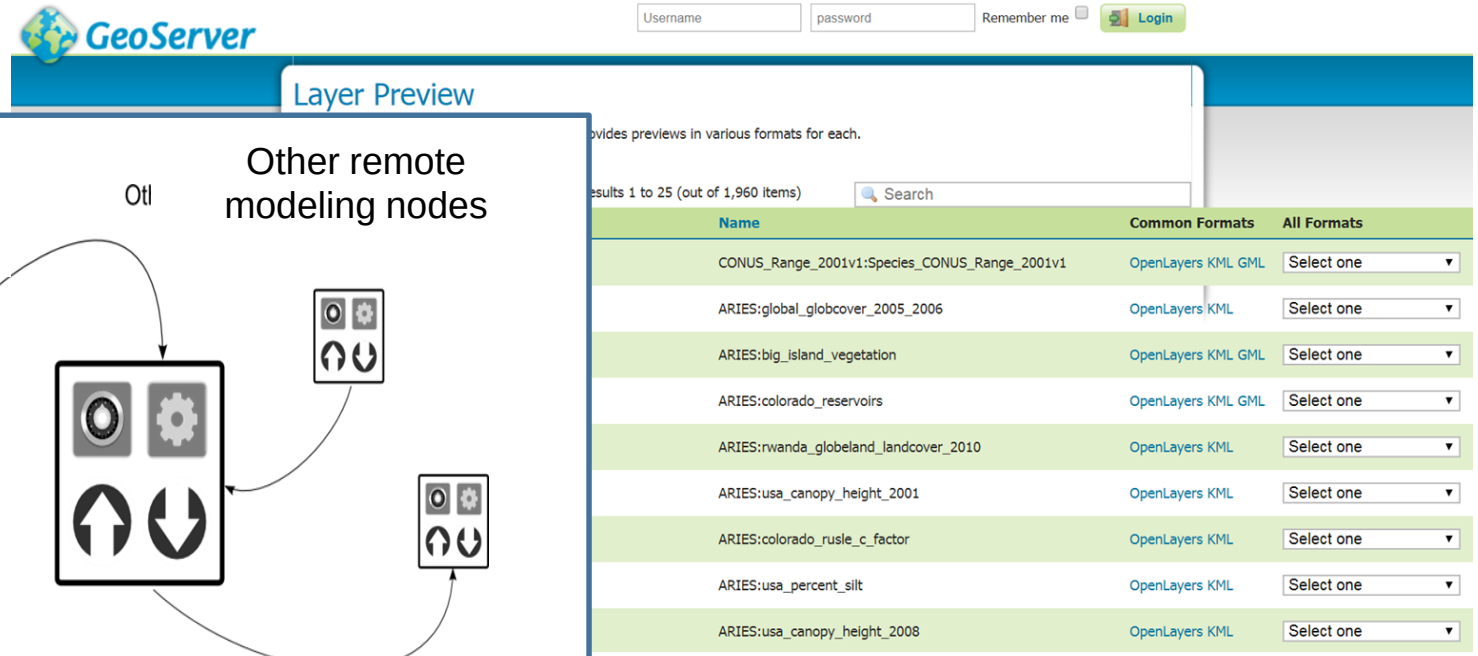


- Archival work is *up front*, after that it's usable by anyone

Cloud-based data and models: Toward context-aware modeling

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





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

Set context: Runoff, Big Island, Hawaii

Query available models

Found 2 models: Wet & dry tropics

IF annual precipitation < 2,500 mm RUN MODEL A, OTHERWISE RUN MODEL B

Find & query annual precipitation

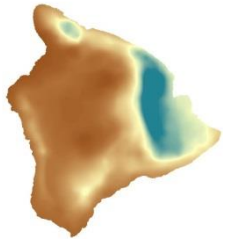
Identify data needs to run each model (DEM, precipitation, soils, vegetation, etc.)

Identified 3 digital elevation models (DEMs)

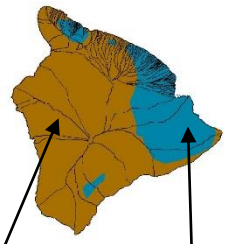
Select appropriate datasets to populate models; run models

Integrated modeling results & provide provenance diagram

Raw precipitation data



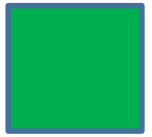
Reclassified precipitation data



Run Model A

Run Model B

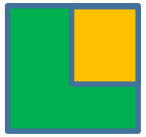
DEM 1: Global DEM, 90 m resolution, reliability score = 70



DEM 2: State DEM, 10 m resolution, reliability score = 85

DEM 3: Study area DEM, 5 m resolution, reliability score = 90

Data integration

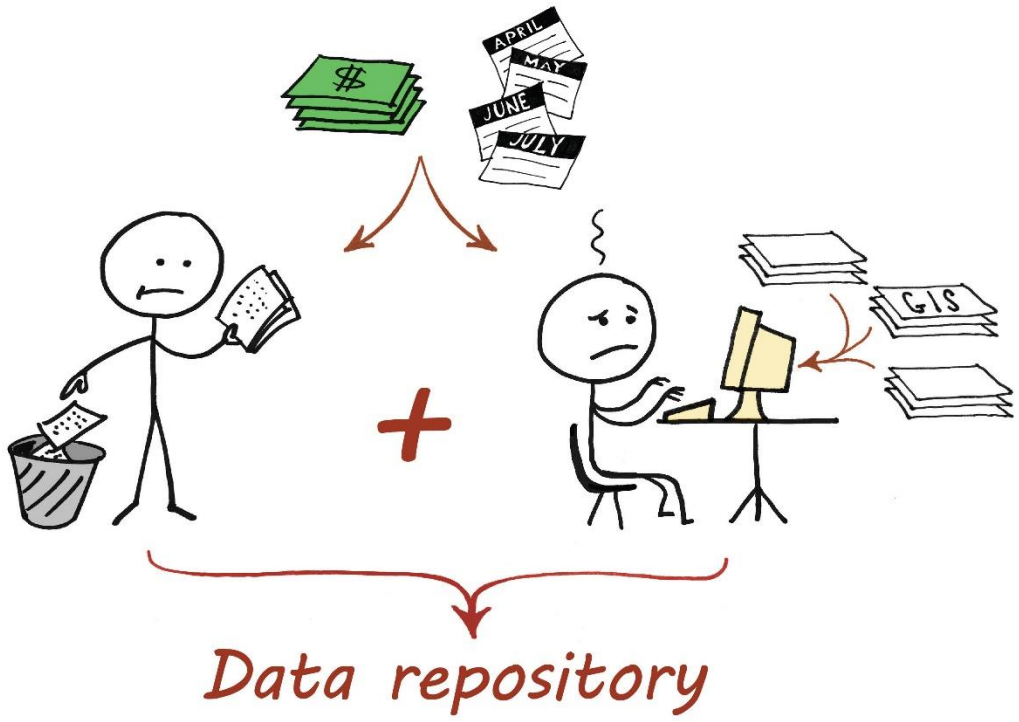


What can we do to make this everyday best practice?

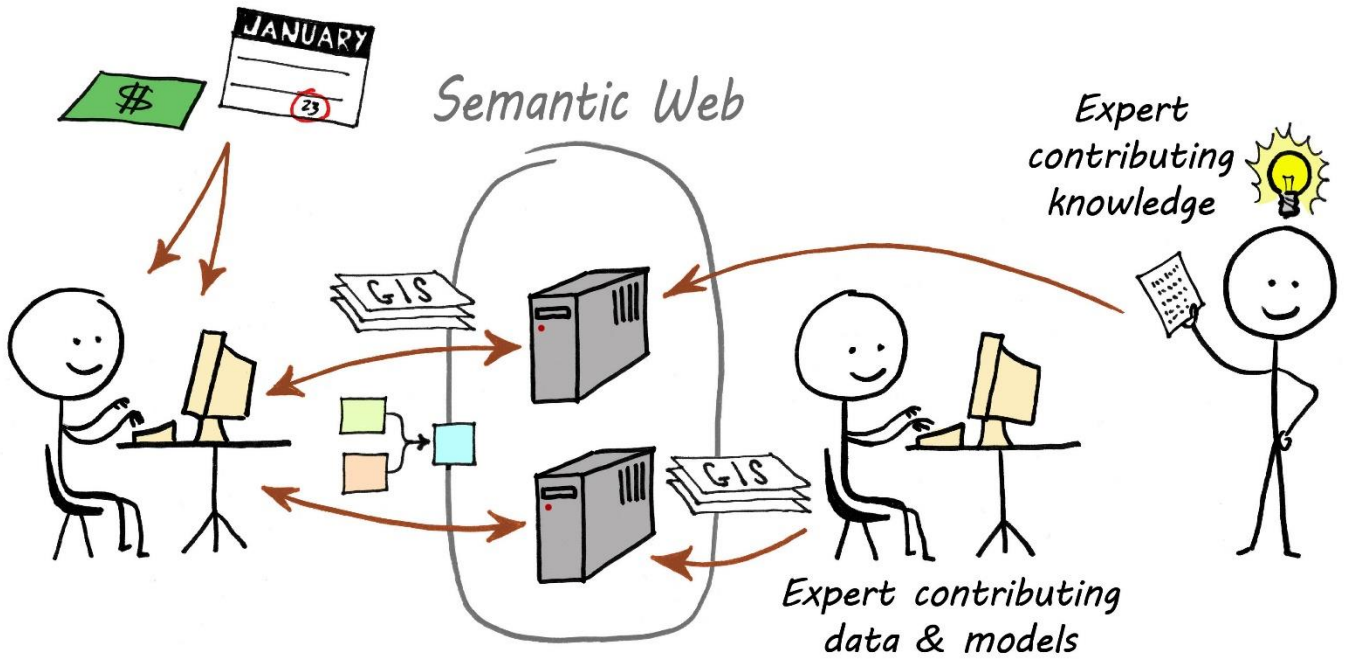
Individual scientists & practitioners can:
1. **Understand** what big data is and isn't
2. **Get comfortable** with use of data-intensive science/inductive modeling when it's appropriate
3. **Collaborate** on and use semantics
4. **Share, serve, annotate data** to facilitate cloud-based semantic modeling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

Streamlining science through interoperability & the semantic web



Status quo



AI & semantic web-supported collaborative modeling

Thanks!

To read more, check out:

http://aries.integratedmodelling.org/?p=1458