An interoperability strategy for ecosystem accounting data & models

Ferdinando Villa, Ken Bagstad, Stefano Balbi, Alessio Bulckaen, Alessandra Alfieri, Bram Edens, William Speller

Forum of Experts on SEEA EA
December 2021

Interoperability:

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

A core challenge to the global SEEA community

Beyond the state of the practice: common goals & standards

Syntactic interoperability:
Use of compatible data formats and communication protocols.
Low bar, more limited advantages

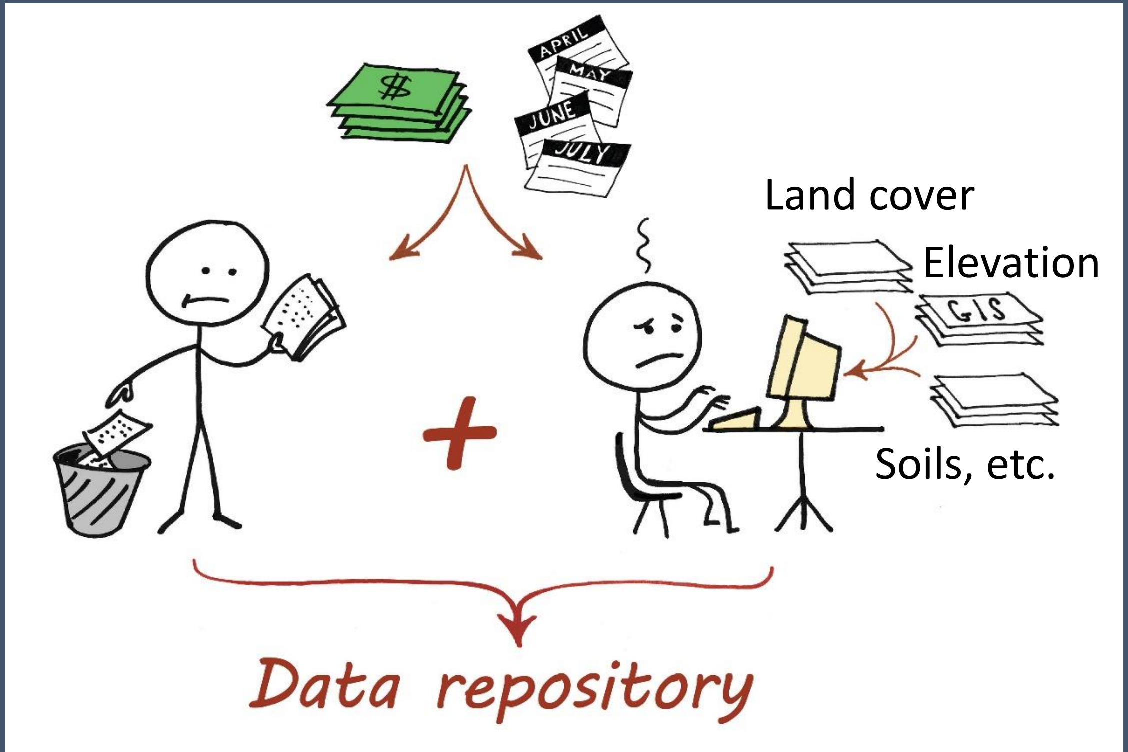
<

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.

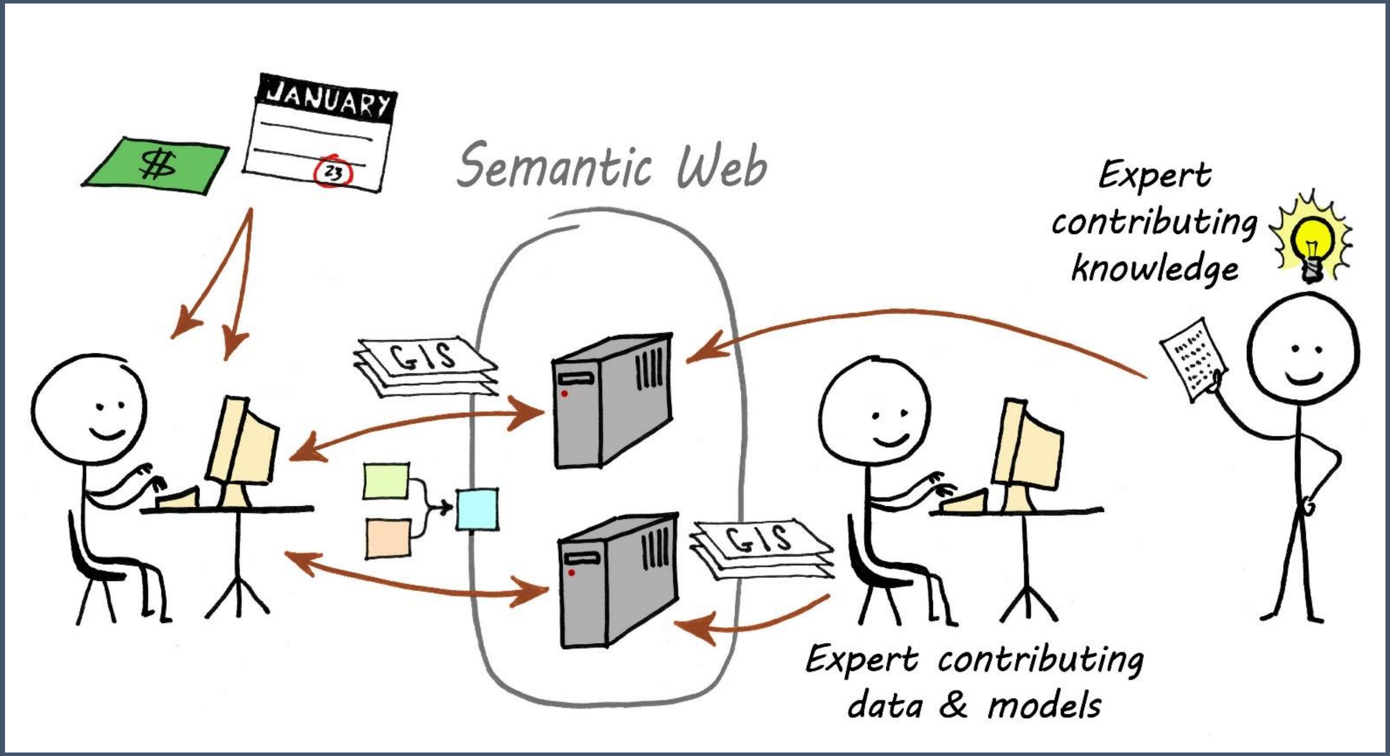
Heiler 1995

Practical example: Sediment retention accounting

Revised Universal Soil Loss Equation (RUSLE): commonly used in InVEST, LUCI, ARIES, and one-off modeling applications



Status quo



Semantic interoperability in ARIES for SEEA

Given the urgency of scaling up SEEA, interoperability is a powerful tool to do so *as a community*

Key building blocks for 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.

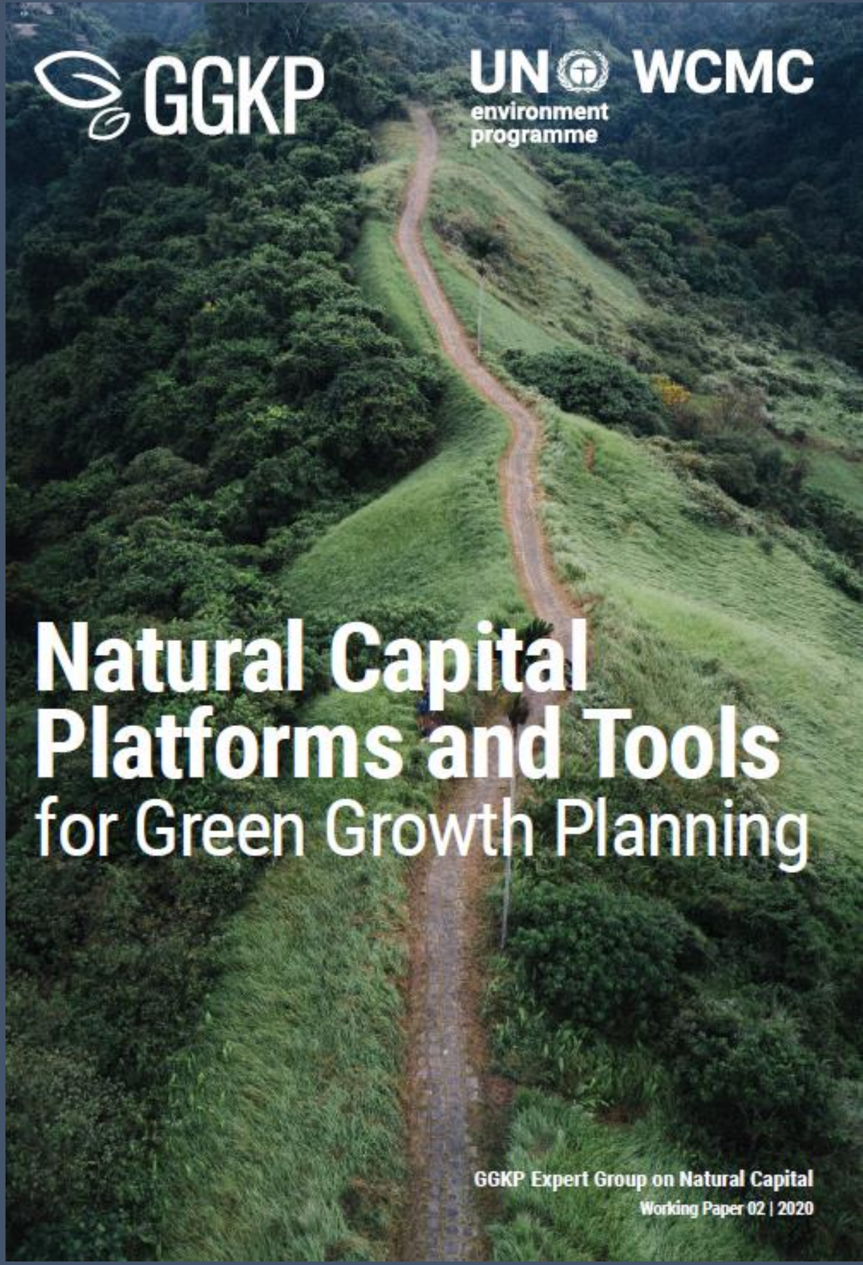
Specify appropriate conditions for safe reuse of your models.

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 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.

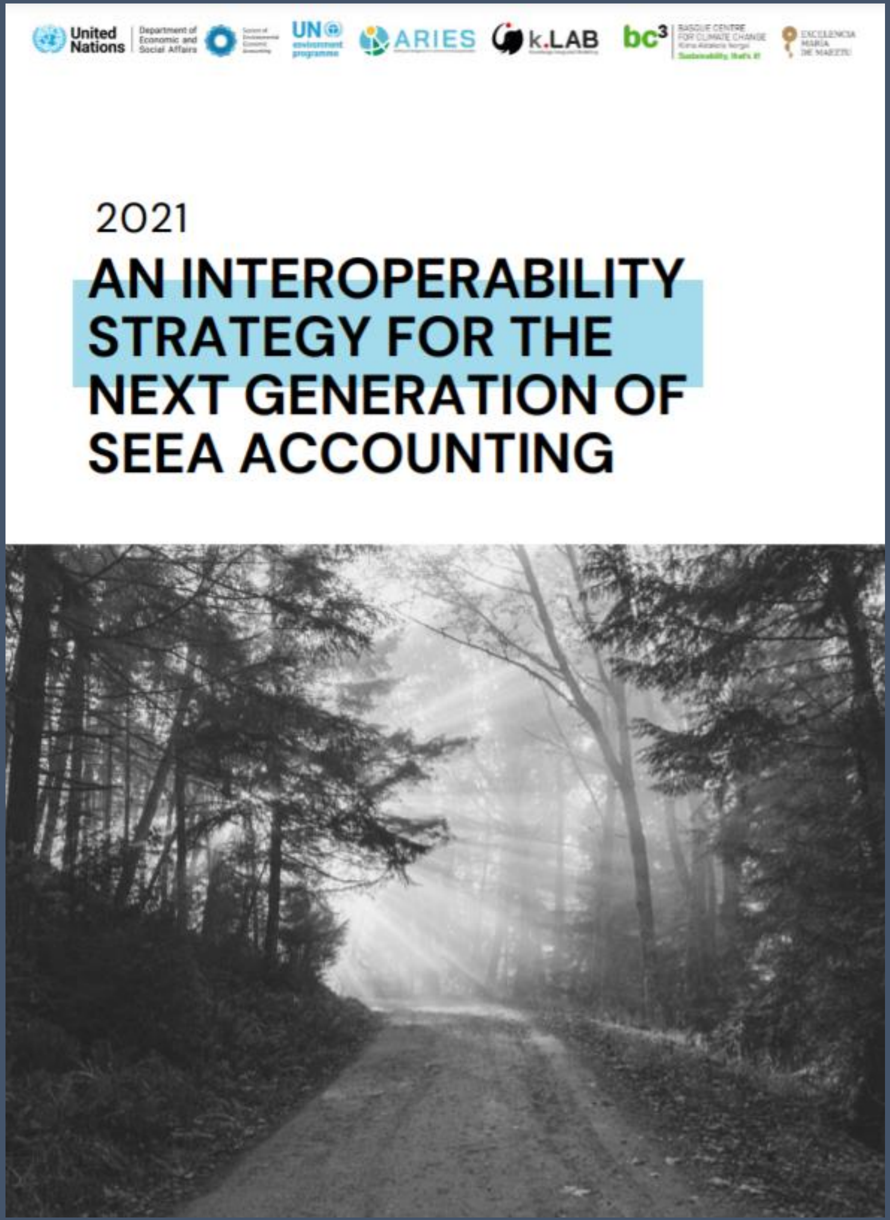


EARTH OBSERVATIONS FOR ECOSYSTEM ACCOUNTING

SEEA interoperability strategy

1. Current state of interoperability & vision for the future
2. Roles & responsibilities (data providers, modelers, institutions incl. NSOs)
3. Implementing the strategy (pilot testing, engaging key stakeholders, governance, training/capacity building)
4. Conclusions

https://aries.integratedmodelling.org/aries-releases-a-strategy-to-scale-up-knowledge-sharing-for-better-informed-policymaking/



Roles of key stakeholders: Data providers (NSOs, science agencies, academic scientists)

1. Expose & maintain key spatial datasets as Open Geospatial Consortium services using networked infrastructure - hosted independently, through the U.N. Global Platform, or other networks explicitly designed for semantic interoperability
2. Use standard coordinate resource systems (projections) that enable on-the-fly reprojection
3. Use open, widely available standards & complete, correct, semantically meaningful metadata; provide an Application Programming Interface (API)
4. Produce Uniform Resource Name (URN)-specified resources from each dataset & publish to a networked node to enable later semantic annotation
5. Use & collaboratively develop common ontologies & vocabularies (allows people & computers to know when data & model components are interchangeable)
6. Identify a point of contact from each institution to follow semantics development & be responsible for their consistent use
7. As a community, gradually move the semantic annotation task to data producers (requires best practice documents, handbooks for specific problem areas, ad-hoc tooling)

Roles of key stakeholders:
Data providers (NSOs, science agencies, academic scientists)

1. Expose & maintain key spatial datasets as Open Geospatial Consortium services using networked infrastructure - hosted independently, through the U.N. Global Platform, or other networks exp
2. Use standard coord … on-the-fly reprojection
3. Use open, widely a … ally meaningful metadata; provide
4. Produce Uniform R … ch dataset & publish to a networked node to enable later semantic annotation
5. Use & collaboratively develop common ontologies & vocabularies (allows people & computers to know when data & model components are interchangeable)
6. Identify a point of contact from each institution to follow semantics development & be responsible for their consistent use
7. As a community, gradually move the semantic annotation task to data producers (requires best practice documents, handbooks for specific problem areas, ad-hoc tooling)

Machine actionable

+

Semantically meaningful

Roles of key stakeholders: Modelers

1. Adopt design principles & guidelines for independently produced, interoperable model projects using distributed version control software
2. Use a modular, non-monolithic model design process to facilitate more interconnected models
3. Learn the importance of tracking provenance of official products; annotate data & models for more informative provenance
4. As a community, develop strategies & incentives to overcome the status quo of noninteroperable model development

Roles of key stakeholders:
NSOs & other institutions

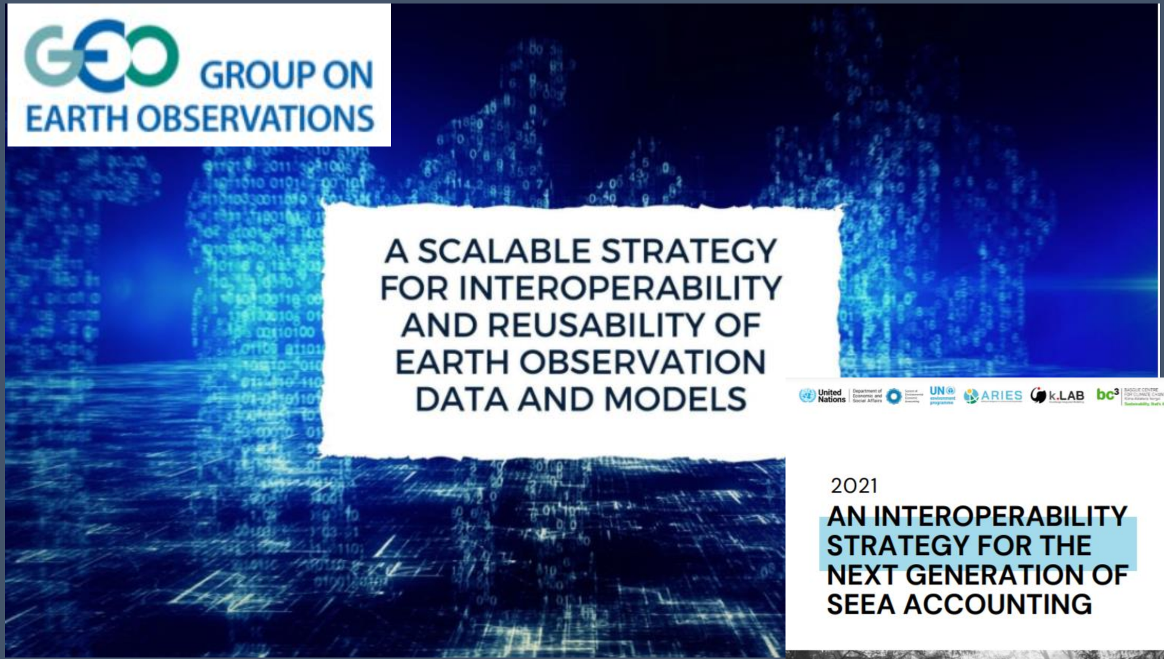
1. Maintaining interoperable resources over time is fundamental (incentives needed)
2. Institutions may formally host data and models; can make results available through APIs or an ARIES/k.LAB node (offers ways to access data and models through both ARIES and other approaches)
3. Institutions doing so require software, hardware, and personnel needs (technical support available)
4. A fully interoperable data & model system moves away from a centralized paradigm to a community of hosting members, within a peer-to-peer system (more stable, flexible, powerful network, each member having full ownership & control of critical data and models & sharing as widely as appropriate)

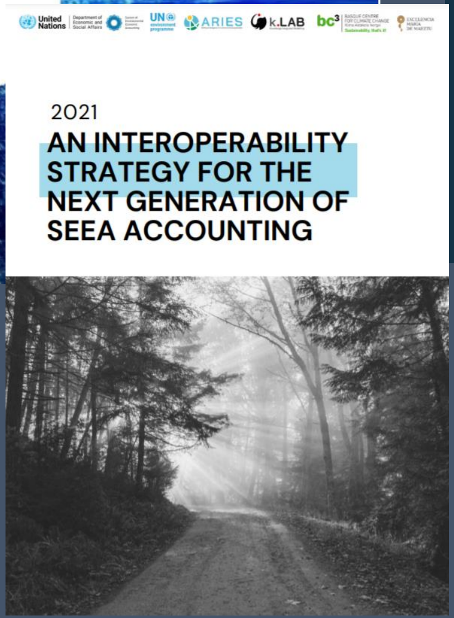
Next steps for interoperable SEEA data & models

1. Pilot testing
2. Engaging key stakeholders (your organizations & others)
3. Governance
4. Training & capacity building

With a *shared vision,*
which uses *common standards*
to *serve the SEEA community's needs*

Interoperability *must address the human element:* User-friendly, equitable, community endorsed

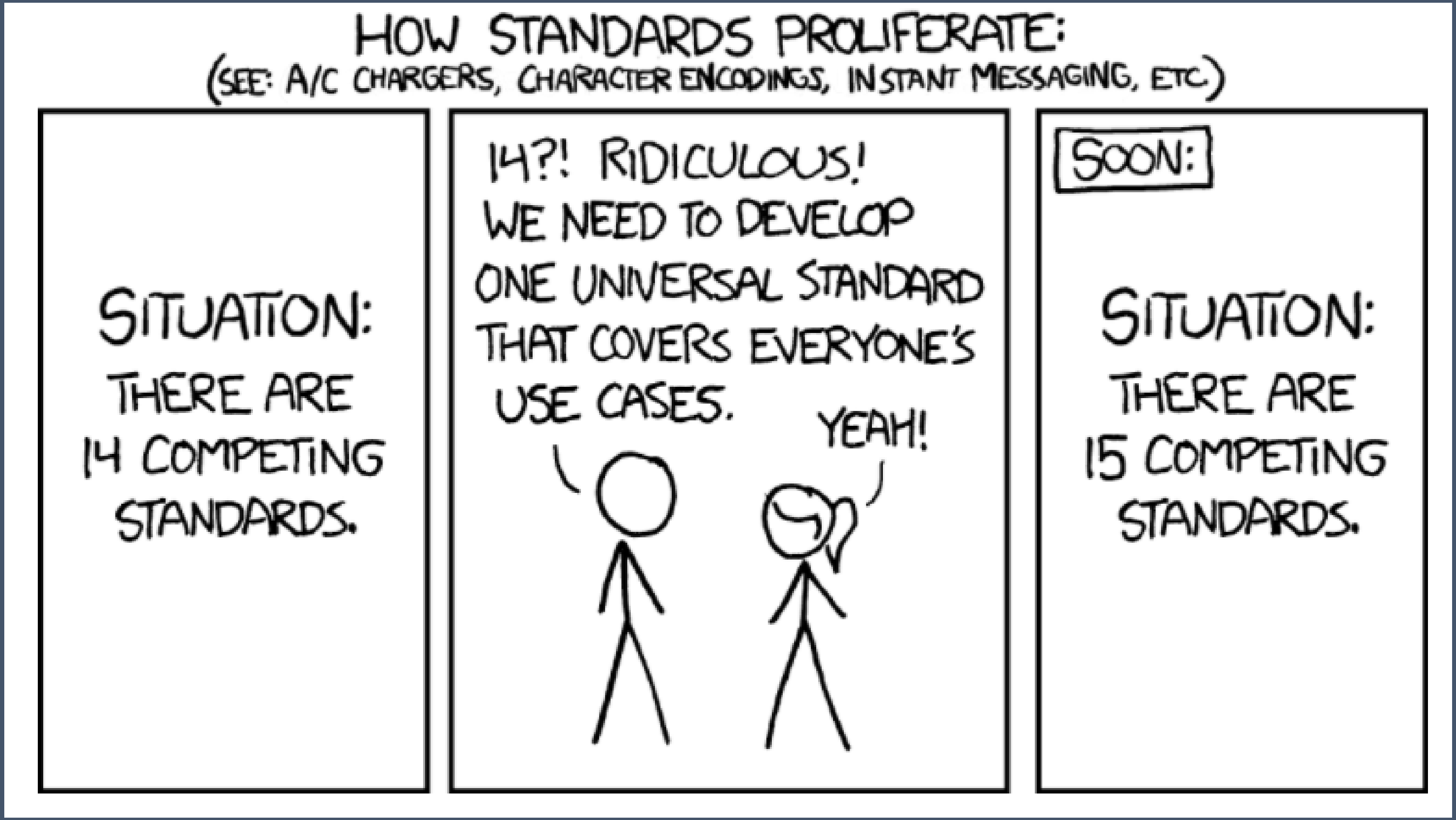




https://www.earthobservations.org/geo_blog_obs.php?id=527

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

Common ground is needed!



https://xkcd.com/927/

Interoperability: Common ground

1. Of the FAIR Principles, Interoperability and Reusability are more difficult to achieve than Findability and Accessibility
2. Interoperability is critical to many science & science-policy goals
3. Open data alone ≠ *machine accessible* data; Open Geospatial Consortium (OGC) standards offer a consensus means of enabling data transfers
4. Semantic interoperability is more powerful than syntactic interoperability
5. Semantics are needed for computers to be able to integrate data and models
   a. Of possible semantic solutions, disciplinary controlled vocabularies are easiest to develop, teach, and use; multidisciplinary ontologies (capable of supporting machine reasoning) are most difficult.
6. Modular model coding is a best practice preferred over monolithic model coding
7. By definition, interoperability implies a need for standards regarding technology & semantics
8. Others?

Interoperability: Points of debate/departure

- Should we aim for semantic or syntactic interoperability? (we believe semantic)
- What proven solutions can be brought immediately to bear on the interoperability problem? (ARIES; others may have different preferences; if so are they ready or many years away?)
- Is the commercial cloud a solution or a barrier to our data interoperability challenges? (we believe it carries risks)
- Others?