

#### **Location Powers: Data Science**

# **Session 2 Summary** Analytics and Representations

Annie Burgess, PhD Earth Science Information Partners | esipfed.org

Nov 13&14, 2019



#### Location Powers: Data Science Session 2: Analytics and Representations

- Moderator: Kumar Navulur, Maxar
- Annie Burgess, ESIP
- Yolanda Gill presentation:
  - Knowledge-Powered Data Science for Integrated Modeling in Geosciences
- Panel on Analytics
  - Todd Mostak, OmniSci
  - Lauren Bennett, Esri Perspective: Spatial data science moves from the avant garde to the mainstream
  - Keith Hare, JTC 1 SQL/GQL convener Perspective: Analytics And Representations in SQL and GQL
  - Hamed Alemohammad, Radiant Earth
  - (Robert Hijmans, UC Davis R-Spatial invited, but could not attend at last minute)

OGC Discussion Group on Analytics



#### Session 2 – 1. What Works



- Adding knowledge enables more-automated and less-manual modeling
- Capturing provenance in models by giving all objects a URI
- Radiant Earth found they needed a community standard for ground referencing. They took the standard to multiple community events, which enabled the creation of a community vetted standard.
- Standards are evolving.

## Session 2 – 2. Open Questions



- Differentiating between prediction and knowledge. When is it okay for it to be a prediction vs. knowledge?
- What *is* ML data vs BIG data?
  - Big data are defined by the Vs, but can be unstructured and unlabeled. ML data must be labeled and structured.
- How to ensure the quality of training data.
- What does it mean to build up the next generation of data science?
  - What should a data scientist know?
  - How should data scientists work with subject matter experts?
- How to maintain training data with a temporal component that changes over time.
- Is there a way to share data select groups (e.g. local governments) with a specialized license? Is there a place for OGC here?

### Session 2 – 3. Next Steps



- Compile training data that matches the diversity of landscapes of interest (e.g. agricultural fields of the world)
- We would like to see more work on the 'raster side' of standards
- We would like to see STAC as an OGC specification
- Convert more EO data to cloud-friendly formats
- Create more domain-specific synthetic training datasets.

