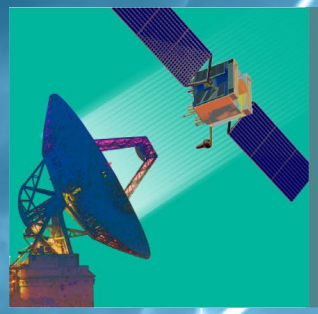
Working Group Session 11D Outbrief



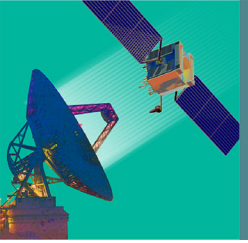
Ground System Architectures Workshop

Opportunities in Data Exploitation

March 2–5, 2020 | Renaissance Los Angeles Airport Hotel

Semantic & Syntactic Consistency – a Critical Enabler for Big Data Analytics

Leads:
Scott Houchin,
The Aerospace Corporation
and David Limbaugh, SUNY Buffalo

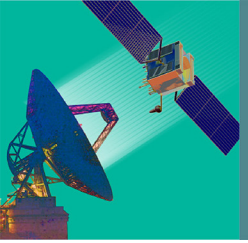


Ground System Architectures Workshop

Session Goals

How do we move toward semantic & syntactic consistency across disparate data sources

- Semantic consistency – The objective: assuring meaning
- Semantic consistency – The value and path forward
- Semantic consistency – The means for representing knowledge
- Semantic & syntactic consistency – the prudent use and development of standards for unambiguous sharing of knowledge



Ground System Architectures Workshop

WG Member Goals & Key Questions

WG members were specifically interested in

- Dealing with the mountains of data, access to the data
- Scraping big archives from multiple data sources and trying to combine them
- Semantics and ontologies
- Data characterization and sharing
- How others develop ontologies

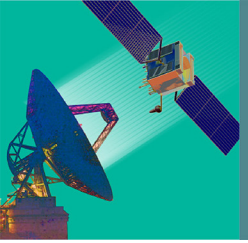
Key questions

- How can we make people agree about the meaning of a term?
- How do we guarantee the same semantic impact of a term whether across users (both machine and human)?
- How to make data discoverable even when the need for the data is not known?
- How can we design data so that one set of tools can easily use data from disparate sources without continual recoding

WG members

Scott Houchin – Aerospace
David Limbaugh – University of Buffalo
Chris O'Hare – JHU APL
Sonia Henry – Aerospace
Scott Bell – TRACLabs
Lynn Roggla – GEOST
Charles Conrad – GEOST
Matt Jacobs – GEOST
Jonathan Dingwall – Qwaltec
Patrick Douglas – Raytheon
Kyle Tunis – Raytheon BBN

Working Group Session 11D

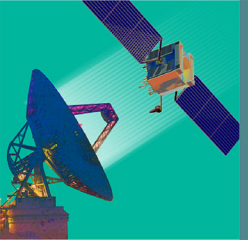


Ground System Architectures Workshop

Key Points

Semantic consistency

- The meanings of terms used to tag data should be consistent a priori.
- We shouldn't rely on machine learning tools to create consistency after the fact.
- We should avoid forcing disparate data into the same data set unless we have to.
- We need groups to supply clear and accessible materials to help non-experts build and use ontologies
- We need to make standardized ontologies more clearly accessible so that they are actually used.

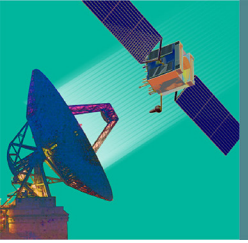


Ground System Architectures Workshop

Key Points

Semantic & syntactic consistency – the prudent use and development of standards for unambiguous sharing of knowledge

- Discussion seeded using a briefing from Scott Houchin on a paradigm for the design and documentation of data formats and standards focused on building blocks
- Our data standards should make the easy things easy and the hard things possible
  - *It should be to meet the immediate need*
  - *It should be easy to extend the data to meet reasonable future needs without breaking compatibility*
- But economic forces sometimes get in the way!
  - *Users ask for data based on their preconceived assumptions, not on their actual needs*
  - *We need to help the users understand what they really want out of the data, then design the data to put those real needs front and center*



Ground System Architectures Workshop

Conclusions

- In an age where our data sets are growing, in size and scope, at enormous rates, our syntaxes and semantics should be flexible, reusable, and scalable.
- This requires cooperation across agencies and organizations to develop new standards of syntactic and semantic development.

- Semantics and syntax are different facets of the data
  - Raw data structure (syntax)
  - Apply meaning to that raw data (semantics)
  - Use data building blocks in combination (syntax)
  - Apply meaning to the combinations (semantics)

- No one is disputing the need and desire for consistency
- No one disputes the theoretical benefits of ontologies
- But there's lots of skepticism that we could get the DoD/IC onto a single standard

- Next year – We want real mainstream demonstrations

Working Group Session 11D