

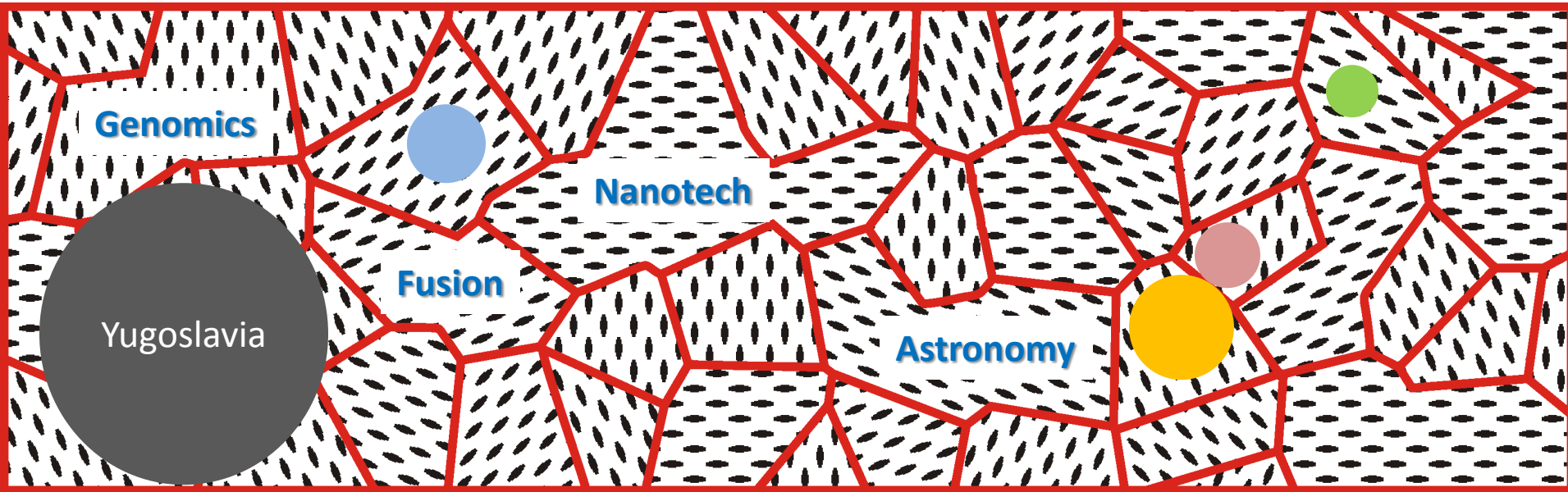
# Workshop Summary and Wrap-Up

Tom Statler, NSF

# With Big Data comes big risk

- Risk of reaching incorrect conclusions
    - through misunderstanding of data
    - through misuse or abuse of data
  - Risk of data investment losing value
  - Risk of data becoming unusable
- 
- Metadata is the essential information needed to minimize these risks.
  - Metadata curation is managing these risks and accepting them where appropriate.

# Aligning metadata across pre-aligned domains?



Standards & practices developed for domain-specific needs are just starting to interact.

The hazard of a top-down unification of standards across domains is that it can appear to lower barriers while being doomed to internal fracturing.

# The sociological problem is as hard as than the technological problem

- Understanding needs of stakeholders
- Sharing can disincentivize competition
- Difficult for researchers to contribute in a new field and yet retain credibility in their home disciplines.
- “Culture eats strategy for breakfast.”
- “It takes internal alignment to get horizontal alignment.”
- Can we measure the value of a dataset?

# Areas of Agreement?

- Active data stewardship/curation adds value and is needed at some level; but we have no funding model to support these people and no way to measure the value compared to, e.g., new research grants.
- We're not looking for exclusively top-down "we're from the government and we're here to help" solutions; but we don't know what the correct balance of grass-roots vs. "middle-out" initiatives is.
- There is a need for easy-to-use tools for metadata creation, improvement, and workflows that incorporate good data practices.

# So what next?

What should federal agencies do (at “zero cost”)?

What should communities do internally?

What should universities/academic researchers do?

What should the private sector do?

Keyword: **DO.** (On-the-job training for everybody!)