Data Citation Breakout Summary - C/D

http://twc.titanpad.com/81
Interesting Data from a Citation Standpoint

- Biological organism classifications - where an organism is in a classification scheme changes as a function of time

- Model runs

- Seismic data organized by network, station, location, channel - granularity, continuous data vs events

- Nimbus 2 satellite imagery on film - concluded that any digitized and cleaned up version of the data would be a new data set

- Data and data subsets generated on the fly (e.g., reformatted, reprojected, visualizations, etc.)

- Aggregated data that is distributed by somebody other than the original archive
Citation issues

• Citation content probably variable depending on purpose:
  • To give credit where credit is due
  • To establish data authority
  • For replication purposes

• Journal acceptance and standards

• Handling data composites - chain of reference (map example)

• Handling static data (e.g., on CD) vs dynamic databases

• Dealing with data subsets and other on-the-fly products (e.g., OPeNDAP vs clients like IDV)
More Citation issues

- What exactly is being cited? The metadata record? The data itself?
- Journal impact factors and citation indices (data journals)
- What about aggregators and distributors?
- Mechanisms for making users aware of the citation and providing tools to generate or access a citation
  - Tools and standards not always set up to handle citations (e.g., OGC services)
- Citation granularity - need best practices
- Carrots for PI's and others
- This whole scientific equivalence issue
Who needs education?

• Data creators (i.e., PI's) especially if they can persuade their colleagues to start too
  • PI's would be able to track down exactly who used their data and how

• Casual data users? reporters and such?

• Educators looking to use data in the classroom (get kids while they are young)

• Need to work with societies publication committees and and publishers (ESIP is the proper group to move this forward)