

Paul Constantine  
Associate dean for Research and Instructional Services  
University of Washington Libraries

Within the context of New Partnerships here are what I see as key issues facing us.

1. Long-term preservation
  - Appraisal
    - There is simply too much data to preserve everything over the long term; we need to devise ways to determine exactly what data is worthy of long-term preservation.
    - Can we develop an understanding of how data might be re-purposed in ways totally unpredicted by its original creator/gatherer/researcher?
    - What legal/ethical constraints encourage or discourage the long-term preservation of particular datasets?
2. Management
  - Who is best positioned to manage long term preservation?
    - Scientists are often ready to move onto to their next project and don't necessarily see the value/need or have the resources to preserve their data over time.
    - Data sitting on a scholar's computer are not easily accessible to other researchers wishing to use and/or repurpose it.
    - Campus Computer Centers are not always funded or equipped to preserve data for the long-term.
    - Neither researchers nor computer centers are especially used to creating metadata schema and assigning metadata in ways to make datasets more easily discoverable.
    - Libraries, while skilled at creating metadata schema and assigning metadata in ways to make datasets more easily discoverable, are generally not funded or equipped to preserve digital data for the long-term.
3. Curation of Scientific Data in Digital Form
  - I see curation in many ways as the overall "thing." Long-term preservation and managements are components of data curation.
  - Funding
    - Many libraries and computer centers are not funded to provide data curation
  - Participation
    - Some researchers need to be convinced of the importance of data curation and the long-term preservation of their datasets.
    - Convincing researchers to share their datasets.
  - Intellectual Property Issues
    - Have the data been copyrights or their use patented?
    - Are the data licensable?
    - Who owns the data?
    - What restrictions have they imposed?
  - Preservation and Access