Mark Sandler
Director, Center for Library Initiatives
CIC
Champaign, IL

SOME KEY ISSUES FOR CONSIDERATION

Archiving

Archivists in academic settings often comment on the lack of interest that scholars show in the notes and working files of their colleagues. Likewise, in the social sciences as well as other disciplinary clusters, there seems to be more reward in gathering new data than building on top of the sources of others. Nonetheless, we know this to be mixed, and ICPSR and other data centers probably have a sophisticated understanding of the kinds of data that does and doesn't support new and important research. We should be tapping into that knowledge to establish guidelines for archiving datasets, both at the level of which datasets should or shouldn't be saved, as well as which should be saved as bit streams with minimal investment as opposed to those worthy of being refreshed, migrated forward and kept readily accessible to subsequent generations of researchers.

Aggregation

Aggregating data is efficient in terms of storage and management, and efficient as well in terms of retrieval. Perhaps disciplinary data farms should be developed, or perhaps this needs to be approached by funding agency or by academic institutions or collectives of academic institutions and research centers. A key to successful aggregation of data will be the emergence of standards around defining variables and achieving a degree of consensus about data gathering techniques that will permit greater comparability across studies. I understand that some differences in survey design and research methodology represent advances in a discipline, but I also understand that far too often these shifts are less about "progress" than idiosyncratic deviations masquerading as a research advance.

In the world of text archives, there is increasing emphasis on standards, keyboarding and scanning guidelines, DTDs, and substantive metadata, all of which further content integration and system interoperability. The underlying theory here is that data gathered for a particular study or purpose might be more valuable as part of a larger whole than if self-contained and self-referential.

Transparency

However it's done, data (especially data gathered with the assistance of public funding) should be more broadly available for public scrutiny and further analysis. I think we all respect the right of a researcher to have the time required for careful analysis of data he or she has gathered before it is opened up to the world. On the other hand, closing down access to useful data for many years on the off chance that the researcher will someday be inclined to return to the dataset seems selfish and not in the best interest of advancing scholarship. As with so many other issues in the underlying social relations of academe, we need to become clearer as a community about the social responsibility of scholars to engage in dialogue with the larger society.