Data Center Models and Impact on Scientific Research Communities Christopher Cieri Linguistic Data Consortium, University of Pennsylvania

The Linguistic Data Consortium (LDC) has served, for more than two decades, as a center for the creation, distribution and archiving of language-related data. LDC employs a Consortium model in which members contribute to via fees and data donations and receive in turn ongoing access to a repository of their collected contributions whose value is several order of magnitude greater. This model, which might be described as stubbornly practical, has survived great economic upheaval and tidal shifts in attitude toward digital data -- but not without change. This presentation will discuss the evolution of the LDC model in comparison with models adopted by other data centers. It will sketch the original plan, the decisions to expand into data collection and annotation, tool building and the distribution of specifications. It will outline and partially quantify the impact the Consortium has had on language related research and technology development. Finally it will discuss the new directions that LDC is pursuing in x-sourcing, web service grids, cloud distribution, and alternate incentives for contributors.