

## Preliminary Thoughts for Your Considerations

It is increasingly commonplace that over the course of the research lifecycle and stages of a scholarly work, digital information is flowing into and out of the workflow at almost every point. This appears to be the case in nearly all academic topic domains. We will take “scholarly communication” to include both human <-> human dialog in all of its forms, and data and instructions that are transferred to and from computational and data infrastructures. The Monday afternoon and Tuesday break out sessions will be in part devoted to identifying distinctive features of scholarly workflow models in the natural and social sciences and the humanities in terms of the nature and level of information flow associated with each stage in a workflow model. Information flow profiles can be strikingly different across and within disciplinary domains. For example, eScience applications rely heavily upon information transfer between computational systems and data infrastructures. Humanities scholarship relies on discursive processes, construction of meaningful arguments from factual instances and querying of linked institutional repositories containing highly diverse information objects with complex representations. Even within sub-domains of certain disciplines, such as those in the social sciences, the differences can be great depending on the methodologies employed (qualitative versus quantitative). The number of possible combinations of topic/methods/tools from which to choose is daunting, but idealized models can serve as a basis for our purposes. (The Digital Commons Network provides a very good visualization of domains and sub-domains as part of its search interface: <http://network.bepress.com/>)

By examining and documenting information flows across each stage of a research activity it will become possible to capture a more complete record of the work. And this is a primary goal. Simply publishing an article that summarizes a research project in a journal is widely regarded as insufficient. Many journals, including Nature and Science require researchers to make data and materials available to peers or others in order to validate the work. Publishing an article with links to datasets and other resource materials in open access repositories is another major step forward, particularly if the data has been prepared for reuse by other researchers. But even this does not fully tell the story. What is needed are very rich document models that include the article, a full description of the work, datasets and sources, technical details, annotation guidelines, links to related work, metadata of all sorts and features that invite other researchers to build on, extend and apply the results in new ways in new application areas.

However, how to accomplish this is difficult to conceptualize given the multitude of entities, processes and actors involved, all of which are dynamic in nature, broad in scope, of immense scale and having great diversity. Conjoined to the modern research lifecycle and scholarly workflows are digital data lifecycles that change as the data objects are given new features, representations and functional primitives. Within the digital data lifecycle are digital curation lifecycles. Hierarchies of

lifecycles emerge regularly and simplified abstract representations of them proliferate.

Another primary goal is to identify changes to current practices that have the potential to increase scholarly research capability and productivity in the near term and at the same time, identify attributes of a scholarly communications infrastructure and how this can be realized. Ideally such an infrastructure will result in beneficial effects on the pace and productivity of the overall global scholarly enterprise. This ambition is increasingly being viewed as tenable as progress is made and credence given to semantic web technologies and robust linked open data infrastructures grow and take hold in the community. It is also important to nurture practices and create tools that lead to semantic annotation of digital content of all sorts associated with scholarly work. These should be designed such that the efforts associated with laborious tasks are distributed equitably across stakeholders and service providers. Finally, we seek means for continuing the rich dialog that has brought us to this point and hope to use this meeting to help guide the way forward.

Steve Griffin  
January 5, 2013