New Collaborative Relationships:
The Role of Academic Libraries in the Digital Data Universe

Position paper, Rick Luce, Emory, September 2006

Many challenges confront the preservation and access to digital scholarly information, as well the development of new capabilities which extend beyond the analog print paradigm. At the heart of these challenges lie infrastructure issues that surround the requirements for interoperability.

1. **Data Model Interoperability -** beyond simple web portal access and harvesting static repositories, raises issues related to defining (and implementing) the data model to utilize for digital objects, which must be commonly represented across heterogeneous, non-static repositories. The data model is important as a higher level of abstraction that must persist over time and support the following:
    - Abstraction for digital objects, required so digital objects can be seen as an instance of the class defined by the data model, and provision for a level of abstraction which persists over time regardless of evolution of changing technologies and formats.
    - Definitions for roles and quality assurance pertaining to the creation and maintenance of metadata, both man and machine generated;
    - Quality assurance pertaining to the curatorial role of automated datasets;
    - Rights, from confidentiality to DRM
    - Sustainability (economic, social, organizational, technological)

2. **Repository Interoperability –** enabling new value chains initiated in repositories that are:
    - cross-repository interoperable and federated. Note: repositories may organized by domain, discipline-orientation, institution or organization, type (e.g., dataset, learning object, format), etc., however, they should not be considered static nodes in a communication system merely tasked with archiving digital objects, and making them accessible through discovery interfaces. Rather, these repositories are be part of a loose, global federation of repositories, and scholarly communication itself is regarded to be a global workflow (or value chain) across dynamic repositories.
    - Support a set of core services utilized via both machine and human user interfaces;
    - Facilitate emergence of richer cross-repository services .

3. **Ecological Interoperable** – which enables:
    - Persistent communication infrastructure, independent of changing technology, which records and expresses the origin and authority of the unit of scholarly communication;
    - Global and automatically executed workflows and grid-enabled workflows which support use and reuse across scholarly repositories;
    - Data provenance in a heterogeneous networked environment;
    - Distributed interoperable instruments and sensor-based registries;
    - Information filtering which automatically pushes information to the user(s);
    - Emergent forms of social software, collaboration environments and networked based user profiles and network traversal log activity analysis;

Increasingly, value resides in the relationships between papers, their associations, and the supporting data sets and materials. To manage and utilize the potentially rich and complex nodes and connections in a large knowledge system such as the distributed web, system-aided reasoning methods would be useful to suggest relevant knowledge intelligently to the user. As our systems grow more sophisticated, we will see applications that support not just links between authors and papers but relationships between users and information repositories, and users and communities. What is required is a mechanism to enable communication between these relationships that leads to information exchange, adaptation and recombination. A new generation of information-retrieval tools and applications are being designed that will support self-organizing knowledge on distributed networks driven by human interaction to support trans-disciplinary science. Through the use of these new tools, we will derive a shared knowledge structure that is based on users and usage in addition to that provided by author citations. Thus, the aggregated connections that readers make between papers and concepts will provide an alternative conceptualization of a given knowledge space. Such techniques will be coupled with classical search and retrieval methods, and these capabilities have an obvious utility for discovering and supporting evolving knowledge from these networks. The same concepts can be applied to data sets and rich media sources.

This emerging adaptive web will analyse and use the collective behaviour of communities of users, utilizing concepts such as adaptive linking, which facilitates the evolution of knowledge structures based on collective user-behaviour over time, and spreading activation, which uses a memory-recall process model from cognitive psychology. For example, using known keywords to search across distributed open archives, a user would receive recommendations of other conceptually related keywords, relevant articles, data sets and so on, based on semantic proximities linked across a multitude of distributed information resources. At the same time, the knowledge system the user has interacted with can begin to reorganize itself by incorporating feedback from the interaction into its knowledge structure.

From the user perspective, such systems can use adaptive webs as a communication fabric to manage and co-evolve the knowledge traded with communities of members and users. Correspondingly, these new tools and systems will influence the adaptation of the structure and semantics of scientific discourse. Many questions remain unresolved, such as how we evaluate the knowledge structures and representations of such size and complexity.