

Beyond Digital Libraries The Organizational Design of a New Cyberinfrastructure

By Donald J. Waters

The curation and communication of information are vital to the development of scholarship generally and essential to the health and progress of specific discipline-based teaching and research. Moreover, just as the accumulation and dissemination of primary sources of data and secondary sources of interpretation and analysis serve as the foundation for knowledge building in the academy, sound principles of information management and communication are essential for successful decision-making in a wide range of business, governmental, military, and other pursuits. Over the last several decades, not only has the growing application of computers and networking demonstrably improved the quality, lowered the costs, and speeded up the work of information curation and communication in many areas, but also work with digital information has opened entirely new perspectives and made advances in knowledge possible in the sciences, social sciences, and humanities that heretofore have been difficult, even unthinkable. These radical, and unanticipated changes have led to cogent and persuasive calls for substantial, long-term programs of investments in research and engineering that would foster and support the development of “ubiquitous, comprehensive digital environments” that are “interactive and functionally complete.”¹

The technical research agenda for creating such a “cyberinfrastructure” is ambitious. Breakthroughs are needed in a broad range of areas, including:

- Robotic and other methods for capturing data rapidly, accurately, in large volumes, and in a variety of formats;
- Facilities for representing and traversing the multi-lingual and multi-layered semantic dimensions of those data;
- Simulation, visualization, and other tools for exploring, comparing, analyzing, and synthesizing information;
- Collaborative filtering, recommendation systems, and other innovative techniques that improve and personalize precision and recall; and
- Systems for ensuring persistence of data and software systems.

The purpose of this document, however, is not to justify the need for and define a program of research to stimulate breakthroughs in these technical areas. Rather, it is to argue that creating a national and international cyberinfrastructure will equally require explicit and ambitious attention to organizational design and other social issues, which

¹ Atkins, Daniel E., et. al., *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure*, January 2003, p. E2. Available online at http://www.communitytechnology.org/nsf_ci_report/. See also Graves, Sara, Craig A. Knoblock, and Larry Lannom, “Technology Requirements for Information Management,” *RIACS Technical Report 2.07*, November 2002, available online at <Http://www.riacs.edu/trs/>.

must also be advanced through a rigorous and thoughtful program of research and development.

The Need

As NSF and the scientific communities contemplate the creation of a cyberinfrastructure, the need for systematic attention to questions of organizational design appears in at least three areas. First, the research agendas for the digital libraries and related programs have regularly made assumptions about the organization of various social factors. The RIACS report on information management, for example, characterizes use, privacy, security, and usability as “attributes of interoperation,” interpreting these factors largely as independent variables which define technology requirements.² These key social factors, however, are themselves highly variable, depending at least in part on the type, purpose, and structure of the organization in which information and technology users are embedded. Because so much technical research in this field depends logically on assumptions about the social organization of information, the validity of the research—and its potential for further development and implementation—depends on much more careful investigation than has been achieved to date of these organizational assumptions, including those about business models and modes of operation.

Second, the Atkins report on cyberinfrastructure wisely distinguishes among the processes of research, development, and operations as essential for promoting and sustaining the software and hardware products of an advanced cyberinfrastructure program (ACP). These three broad classes of activity, of course, are interrelated and ideally feed back options and requirements to one another. Such feedback would be an essential component of the ongoing vitality of the program, but it is not sufficient. To administer and help sustain ACP, the report further recommends both an internal organization within NSF, and a community-based structure of centers. These centers would foster development activities as well as user support and other operations at both generic and disciplinary levels. Although the recommendations for organizational change at NSF are quite specific, the design of the community-based centers is left unspecified, and will presumably vary, in part, with the needs of the academic communities they are intended to serve. However, if the centers are to help sustain and support the products of ACP, they must be designed in such a way that they operate in a business-like fashion and can sustain themselves, eventually independent of NSF support. Such design should not fail to be informed by current expert understandings and additional targeted research that would indicate how certain types of mission, leadership, governance, organizational structure, legal arrangements for intellectual property, and financing, especially in the context of public goods economics, could contribute to—or undermine—the success of such centers.

Third, many digital library and related research projects depend on the use or creation of substantial databases of content from one or more subject domains.

² Graves, et al., *op. cit.*, pp. 18-19.

Moreover, subject-based research within specific academic disciplines also often yields significant collections of data and other content that may themselves be valuable outcomes of the funded research, and essential ingredients of an emerging cyberinfrastructure. However, such databases are often created only with narrow research uses in mind, and efforts to move them into more broadly based, self-sustaining, operational use are often stymied. Just as there needs to be robust, feedback-laden models for software and hardware products to move from research to development and operations, there need to be similar models supporting the life cycle of content in the sciences as well as other domains, especially for those forms of information that are specific to the digital environment and for which new types of organizations are needed because reliance on publishers, libraries and other traditional means of dissemination do not, or cannot suffice. If it is to endure, digital content cannot be an afterthought, but must be actively developed and curated by designated leaders of specific organizations who are charged with data management, responsible for safeguarding property, use, and users, and are fully accountable for their actions in a clearly defined governance structure. In other words, just as the design of the ACP centers would need to be informed by current expert understandings and additional targeted research regarding organizational factors such as mission, leadership, governance, organizational structure, legal arrangements for intellectual property, and financing, so too would the design of organizations responsible for cyberinfrastructure content.

A Suggested Program

A hugely fertile area for research and development to support the curation and communication of scholarly information in the sciences and other disciplines thus falls under the broad topic of organizational design. Electronic resources do not need to be managed within existing organizational structures, but to persist they must be managed within some organizational context, and as the previous section has demonstrated, the emerging cyberinfrastructure presents substantial new challenges in organization and governance. On the one hand, with investment in technology, barriers to entry for the creation and management of digital resources can be lower than they are when the storage of physical items requires large capital investments in physical objects and buildings to house them, but small institutions that want to develop, provide, and manage electronic resources often lack the sophisticated curatorial, legal, financial, and other organizational skills that are necessary. On the other hand, the huge economies of scale that are possible with digital databases are difficult to manage over current institutional boundaries. Clearly, new organizations and organizational models are needed that are sensitive to the dynamics of particular scientific communities, driven by academic mission, and able to sustain themselves over time as integral parts of the broader cyberinfrastructure. To foster the development of appropriate organizations and organizational models, NSF should institute the following programs and features as part of the broader Advanced Cyberinfrastructure Program (ACP):

- 1) Research on organizations and organizational design. There would be two broad objectives to this research. On the one hand, it would identify the organizational

variation within an academic community or set of academic communities that would affect the requirements and parameters for research, development, and operation of new technology funded as part of the ACP. A second objective would be to take various scenarios of research, development, and operation, and explore the advantages and disadvantages for the emerging cyberinfrastructure of different mixes of organizational features.

The research on organizational design should focus on the following organizational variables: types of mission such as commercial and non-profit; types of governance, including membership, board, and partnership models; leadership qualities; structural dimensions, such as size; policy issues, such as privacy, security, and risk management approaches to the ownership and use of intellectual property; and financing options, taking account of the importance of common or public good economics for the emerging cyberinfrastructure. Funded work could employ a mix of empirical case studies and theoretical approaches, and it could be embedded as part of a larger project or conducted as a standalone initiative.

- 2) An apparatus for incubating and supporting new organizations. In order to create a sustainable cyberinfrastructure, the new centers and content-management organizations should have access to a highly specialized organization—or set of organizations—that can provide expert advice on questions of mission, leadership, governance, and general business practices, so that when created, new organizations created or struggling to survive in the cyberinfrastructure have a reasonable chance of operating in a business-like fashion. The supporting organization(s) must not operate in a “cookie-cutter” fashion, but must be sensitive to variations in need among academic communities, as well as to differences in size and trajectories of growth. Ideally, to economize on the costly duplication of services, the supporting organization(s) might also take direct responsibility for providing a set of common services, such as accounting, human resources, board governance, and legal advice, thereby helping to create a family (or families) of efficiently run organizations.

Donald J. Waters