Chapter III

Addressing the Metadata Gap: Ad Hoc Digital Documents in Organizations

Lisa D. Murphy
Louisiana State University, USA

INTRODUCTION

In today’s World Wide Web, Napster, Active Server Page world, distinctions among text, databases, and documents are disappearing both inside and outside a firm’s firewall. This chapter brings out of the shadows a key issue related to document work in organizations: metadata for ad hoc digital documents. Ad hoc digital documents are those created, stored, and used outside the purview of formal text databases or document management systems. When every end user is a potential publisher and every desktop computer a potential server, document management is not just the concern of records retention personnel and information systems professionals. And, researchers and technologists applying their sophisticated automated summarization and full-text searching tools to documents under user control must prove the worth and improve the effectiveness of their tools.

Digital documents are crucial for day-to-day organizational work, but as yet their management is personalized, erratic, unsystematic; both the research on text databases and the practice of document management can benefit from
an improved understanding of this complex context and the roles of metadata in it. We will review empirical studies of document work in organizations and consider both traditional and emerging roles for metadata in ad hoc document work settings. After identifying these roles, we will analyze a metadata standard implemented in HTML (the Dublin Core) for its suitability for supporting metadata use in this context. Directions for practice and research are offered prior to the concluding remarks.

BACKGROUND

In organizations, document workers with a typical networked desktop computer can create complex, multimedia, multi-application, hyperlinked, digital documents with embedded procedures, sophisticated metadata, and automatically generated summaries. They can distribute the documents electronically through e-mail, file sharing, and Web pages, and enforce a variety of controls over the documents’ use, including file attributes, annotation controls, peer-to-peer networking, and passwords. Those with Lotus Notes can intermingle documents, data, and e-mail, revise the structure and relationship of documents over time, and use the software’s sophisticated replication functions to keep distributed copies of these “databases” synchronized. Often document workers can do all this without involving an information systems or document management professional. However, the presence of advanced publishing capabilities does not guarantee good document practice, make appropriate information available to others, or assure compliance with company policies, regulatory requirements, or legal needs.

Beyond the complications of these new and more powerful technologies, today’s ways of working also have implications for document management. Digital documents move quickly, and generally without impediment, around organizations and across hierarchies, and often between them as well (Murphy, 1999). Traditional boundaries of information processing within organizations—task, work group, department, division (e.g., Daft and Macintosh, 1981)—are no barrier to digital documents transmitted as e-mail attachments or as copies put on shared hard drives. Adding to the complexity is the increasing use of nonhierarchical organizational structures such as task forces, virtual teams, and communities of practice (Wegner & Snyder, 2000).

The new ways of organizing work interact with the new technologies. Groupware of various stripes can be used to support group activities ranging from a onetime meeting to a globally distributed virtual team. Computer-based communication tools such as these have been identified as the “back-
A Systematic Approach to Flexible Specifications, Composition, and Restructuring of Workflow Activities
www.igi-global.com/article/systematic-approach-flexible-specifications-composition/3304?camid=4v1a

A Measurement Ontology Generalizable for Emerging Domain Applications on the Semantic Web
www.igi-global.com/chapter/measurement-ontology-generalizable-emerging-domain/4290?camid=4v1a