International Journal of Information Management 30 (2010) 256–264

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Organization of biomedical data for collaborative scientific research: A research information management system

Sahiti Myneni[a,*], Vimla L. Patel[a,b]

[a] Center for Decision Making and Cognition, Department of Biomedical Informatics, Arizona State University, Phoenix, AZ, United States
[b] Department of Basic Medical Sciences, University of Arizona College of Medicine – Phoenix, in Partnership with Arizona State University, Phoenix, AZ, United States

ARTICLE INFO

Article history:

Keywords:
Biomedical data
Bioscience
Information management
Implementation
Collaboration

ABSTRACT

Biomedical researchers often work with massive, detailed and heterogeneous datasets. These datasets raise new challenges of information organization and management for scientific interpretation, as they demand much of the researchers' time and attention. The current study investigated the nature of the problems that researchers face when dealing with such data. Four major problems identified with existing biomedical scientific information management methods were related to data organization, data sharing, collaboration, and publications. Therefore, there is a compelling need to develop an efficient and user-friendly information management system to handle the biomedical research data. This study evaluated the implementation of an information management system, which was introduced as part of the collaborative research to increase scientific productivity in a research laboratory. Laboratory members seemed to exhibit frustration during the implementation process. However, empirical findings revealed that they gained new knowledge and completed specified tasks while working together with the new system. Hence, researchers are urged to persist and persevere when dealing with any new technology, including an information management system in a research laboratory environment.

1. Introduction

Biomedical informatics, an inherently interdisciplinary and integrative field, has great amounts of data ranging from the public health clinical research to the genomic research. Biomedical research when coupled with the high speed processing technologies results in highly detailed data sets (Roos, 2001). With increased emphasis on translational and collaborative research, vast amounts of heterogeneous data are generated. Managing and sharing these data is vital for subsequent analysis in the biomedical domain (Lyons-Weiler, 2005). The following topics on biomedical research information management are the most studied in literature: database architectures, development of ontologies, and data integration techniques (Topaloglou, 2004). While previous researchers have conducted needs assessments of biomedical researchers from a system design perspective (Anderson, Ash, & Hornoch, 2007), few studies have examined the impact of existing laboratory data management practices on bioscience research.

The primary purpose of this current study was to identify and categorize the shortcomings of the existing laboratory data management practices from the perspective of the principal investigators and laboratory members. In addition, the authors emphasized the importance of new technology in response to discovered limitations and analyzed the implementation challenges when such new system was introduced into the laboratory environment.

2. Biomedical research data

Biomedical data can be of various forms drawn from a wide range of sources such as images from CAT and MRI scans; signals from EEG; laboratory data from blood, specimen analysis; and clinical data from patients. Growing barriers between clinical and basic research are making it more difficult to translate newly generated scientific knowledge at the bench into clinical practice at the bedside. With recent National Institute of Health (NIH) priority for translational research, organization of the basic laboratory data and clinical data has become significant (NIH, 2008). In this paper, we primarily focus on basic laboratory data and its management. However, Section 2.2 gives some insight into the nature of clinical data organization.

* Corresponding author at: Center for Decision Making and Cognition, ABC 1, 425 N Fifth Street, Arizona State University, Phoenix, AZ 85004, United States. Tel.: +1 602 827 2559; fax: +1 646 349 4081.
E-mail address: smyneni@asu.edu (S. Myneni).

0268-4012/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijinfomgt.2009.09.005

2.1. Bioscience laboratory data and its organization

Genomic research laboratories are one of the primary data sources in biomedical research. They are data intensive as evidenced by the immense databases generated by the Human Genome project (ConsortiumIHGS, 2001). The data processed in a genomic laboratory range across the DNA sequence, mutation, expression arrays, assays, antibodies, and oligonucleotides to name a few. The challenge of genomic medicine lies in analyzing and integrating these diverse and voluminous data sources to elucidate normal and abnormal physiology (Louie, Mork, Sanchez, Halevy, & Hornoch, 2007). The manner these data are organized in a research laboratory plays a key role in aiding and driving the research coherently. Current laboratory data management methods primarily include handwritten laboratory notebooks, paper files, homegrown small databases and spreadsheet files (Anderson, Lee, et al., 2007). The impact of these techniques on the bioscience laboratory research is discussed in Section 3.2.

2.2. Clinical data and its organization

Like scientific laboratory data, clinical data need to be well organized to generate adequately balanced results in the realm of translational research. Most of the clinical data often appear in free-text form with little or no structure (Schweiger, Hoelzer, Rudolf, Rieger, & Dudeck, 2003). In their raw form, clinical records consist of hundreds of test results, medication and appointment notes. Illegibility of handwritten documents and inability to access data from various clinical sources greatly limit the effectiveness and efficiency of traditional paper based clinical records. Such drawbacks of paper based clinical records triggered the advent of computer-based medical records (Shortliffe & Barnett, 2006). Contrary to traditional paper records, data recorded in electronic medical record (EMR) systems is legible, remotely accessible and better organized because of the structure imposed on the data input (Tang & McDonald, 2006). Electronic medical record systems, however, are not flawless. Studies show that the use of computer-based patient record technology may cause unintended problems such as loss of relevant and critical information (Patel, Arocha, & Kushniruk, 2002).

In summary, handwritten paper files and homegrown databases are usually used for managing the basic research data, while electronic medical records are increasingly used for handling clinical data. Next, the authors will examine the problems that bioscience researchers often face with the current methods of basic laboratory data organization.

3. Current trends and issues of biomedical data management

Like other basic sciences, recent advances in genetics and general laboratory methods have led to a tremendous increase in the amount of research data captured and analyzed by research teams. Unfortunately, existing commercial software and LIMS (Laboratory Information Management Systems) are unable to organize such data collected from modern bioscience research laboratories, and meet individual researcher's needs (Anderson, Lee, et al., 2007). The authors investigated two scientific laboratories (referred to as labs now on) to understand the influence of these data management methods on bioscience research. While six candidate labs were considered, two test labs were selected based on their responsiveness, motivation of the lab's Principal Investigator (referred to as PI now on), and the richness of lab environment in terms of its ability to represent the manifold changes of use of information technology to improve scientific productivity and laboratory researcher's satisfaction in the realm of bioscience research.

3.1. Data collection methods

Ethnographic observations were conducted. A trained researcher unobtrusively observed the activities at different times in the test labs and took observational notes (Van Maanen, 1996). The purpose of ethnographic observations was to understand workflow of the bioscience labs, and gain insight into interaction strategies among lab members in order to guide and improve efficiency of data collection in the next phase. The important concepts identified during the ethnographic phase were used to design web-based questionnaires. Two questionnaires (Q1 and Q2) were used in this study with the lab PIs to understand the information management practices followed in the test labs. Q1 was administered to all six candidate lab PIs during the test lab selection process, while Q2 was given only to the PIs of the two selected test labs. Both the questionnaires included open-ended and closed specific questions. The questionnaires served as a means:

- to gain knowledge about the overall state of labs in terms of (1) magnitude and nature of data handled, and (2) data management techniques,
- to create an account of current data handling and communication practices in the two test labs.

The participants responded to all the questions and based on their questionnaire responses the themes for the semi-structured interviews were framed. Unlike the questionnaire framework, where detailed questions were formulated ahead of time, semi-structured interviews began with more general unstructured questions (Bernard, 2002; Crabtree & Miller, 1992). Semi-structured interviews provided an opportunity to learn more about the laboratory goals and practices. There interviews allowed us to collect detailed descriptions to understand the reasons behind the problems faced by current day bioscience researchers. A number of new questions were generated during these interviews, allowing both the interviewer and interviewee to probe further on a particular issue(s). The four interview areas of interest were laboratory data storage, laboratory data management, queries on stored data, and collaboration. Nine test lab members in different professional roles such as lab manager, computer support specialist, and bench molecular biology investigators were interviewed. These interviews contained rich descriptive accounts of specific team members' roles and activities. All interview data were audio recorded and transcribed for analysis. Beyond in-person interviews, we plan to conduct online semi-structured interviews. This "*mixed mode*" interviewing strategy will be included in the next phase of the study (Meho, 2006). Finally, Google Documents were utilized to record and track the progress of the study. Detailed analysis and discussion of Google Documents is presented in Sections 5 and 6.

3.2. Data analysis

According to the data collected during the initial recruitment survey (Q1), both the test labs have been dealing with human subjects, paper medical records, locally made DNA and RNA proteins, tissue blocks, microscopic slides, radiogram films and a wide range of biomaterials such as oligonucleotides and antibodies. Table 1 presents an overview of the data collected during Q1. As indicated in Table 1, test lab II has been established for 15 years, while test lab I only has been in operation for 7 years. Therefore the magnitude of data handled by test lab II is greater than that of test lab I. For this reason, the evaluation of information management practices