Published in final edited form as:

Comput Methods Programs Biomed. 2016 April ; 126: 160–170. doi:10.1016/j.cmpb.2015.11.001.

Resolving Complex Research Data Management Issues in Biomedical Laboratories: Qualitative Study of an Industry-Academia Collaboration

Sahiti Myneni¹, Vimla L. Patel², G. Steven Bova³, Jian Wang⁴, Christopher F. Ackerman⁵, Cynthia A. Berlinicke⁶, Steve H. Chen⁴, Mikael Lindvall⁵, and Donald J. Zack^{3,6,7}

¹School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX

²New York Academy of Medicine, New York, NY, and Department of Biomedical Informatics, Arizona State University

³Departments of Pathology, Genetic Medicine, Health Sciences Informatics, Oncology, and Urology, Johns Hopkins University School of Medicine, Baltimore, MD

⁴BioFortis Inc., Columbia, MD

⁵Fraunhofer Institute for Experimental Software Engineering, College Park, MD

⁶Wilmer Eye Institute

⁷Institute of Genetic Medicine Johns Hopkins University School of Medicine, Baltimore, MD

Abstract

This paper describes a distributed collaborative effort between industry and academia to systematize data management in an academic biomedical laboratory. Heterogeneous and voluminous nature of research data created in biomedical laboratories make information management difficult and research unproductive. One such collaborative effort was evaluated over a period of four years using data collection methods including ethnographic observations, semi-structured interviews, web-based surveys, progress reports, conference call summaries, and face-to-face group discussions. Data were analyzed using qualitative methods of data analysis to 1) characterize specific problems faced by biomedical researchers with traditional information management practices, 2) identify intervention areas to introduce a new research information management system called Labmatrix, and finally to 3) evaluate and delineate important general collaboration (intervention) characteristics that can optimize outcomes of an implementation

Corresponding Author: Sahiti Myneni, PhD, MSE, 7000 Fannin, Suite 600, Houston, TX 77030, Phone: 713.486.0115, Fax: 646-349-4081, Sahiti.Myneni@uth.tmc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosures: The authors have **DISCLOSURE**

Under a licensing agreement between BioFortis, Inc. and the Johns Hopkins University, Dr. Bova and the University own BioFortis, Inc. stock, which is subject to certain restrictions under University policy. The terms of this arrangement are being managed by the Johns Hopkins University in accordance with its conflict of interest policies.

Steve H. Chen, Yakov Shafranovich, and Jian Wang are employees of BioFortis Inc., maker of the Labmatrix software.

process in biomedical laboratories. Results emphasize the importance of end user perseverance, human-centric interoperability evaluation, and demonstration of return on investment of effort and time of laboratory members and industry personnel for success of implementation process. In addition, there is an intrinsic learning component associated with the implementation process of an information management system. Technology transfer experience in a complex environment such as the biomedical laboratory can be eased with use of information systems that support human and cognitive interoperability. Such informatics features can also contribute to successful collaboration and hopefully to scientific productivity.

Keywords

information management; industry-academia collaboration; cognition; technology transfer; biomedical laboratory; informatics implementation; virtual research environment

1. Introduction

Biomedical research when coupled with the high speed processing technologies results in highly detailed datasets. Bioscience laboratories are inherently data intensive which is evident from the publicly accessible immense databases generated by the Human Genome project [1]. The information to be processed in a biomedical laboratory ranges from DNA sequences, mutation analysis, expression arrays, cell and biochemical assays, animal and reagent inventory ,to name a few . The challenge to genomic medicine is to analyze and integrate these diverse and voluminous data sources to elucidate normal and abnormal physiology [2-4]. In addition to scientific research, academic biomedical laboratories (labs) are expected to use their data and consequent results to support education, knowledge dissemination through publications, and future research questions. However, current traditional informatics approaches used by these labs have helped them little in performing their daily tasks and in fact, jeopardized their productivity [5,6]. Some of the current lab data management methods include handwritten lab notebooks, paper files, homegrown small databases and spreadsheet files. Problems with heterogeneous data, nonintegrated information retrieval and querying, nonadherence to standards, lack of documented laboratory workflows, and usability issues have been reported as major hindrances to efficient information management and research productivity in bioscience labs. The need for informatics solutions to solve the problems with current information management methods used by academic as well as industrial biomedical labs is well-documented [7-9]. Informatics systems that (a) enable collaborative authoring of lab records, (b) integrate and parse multiple biomedical data types, (c) efficient database architectures, and (d) user-centered data visualization have been developed and evaluated [10-15]. With increase in functional complexity in biomedical research, a single system can no longer support the entire lifecycle of a biomedical research project. Hence it is inevitable for lab researchers to use multiple interoperating systems for their research. Modern information management systems should have a common layer of interoperability while providing a spectrum of options that could be used to support individual researcher needs. Comprehensive, yet nuanced customization of information management systems, therefore may be a solution to address informatics issues in bioscience labs.

In this paper, we describe the attempts of our collaboration to identify and reduce barriers to scientific productivity and satisfaction in a specific translational research-oriented biomedical research lab through direct involvement in evaluating and improving data management in research laboratories. Contrary to the focus of our paper, published literature on similar collaborative efforts have provided insights into technical architecture of bioscience system with minor focus on sociotechnical aspects [16], and summative evaluation of an informatics-driven solution rather than implementation process itself [17]. Majority of reported studies on bioscience information management provide an account of system design and demonstration [18-21]. Our study is data-driven and qualitative in nature. Specific objective of this paper is to present the efforts made by one industry-academia collaboration to improve the data management and integration capabilities of an academic biomedical lab through an iterative design process. The lessons learned from this collaborative implementation effort can inform the scientific, academic, and industrial communities of (a) implementation strategies that can help optimize technology transfer in academic biomedical labs, (b) possible pitfalls to be aware of during intervention with a new data management system, and (c) important characteristics that are pre-requisite for every biomedical research collaboration when implementing new information technology in academic labs. The next sections are organized as follows. Section 2 provides a brief overview of our study context, outlining the industry-academia collaboration with a brief background on the nature of this intervention project as well as of the technology that was used to improve management of biomedical research data. Section 3 describes methodological details of the study, and in the remaining sections of the paper, we summarize the results and limitations of the study while providing relevant discussions and conclusions.

2. Study Background

Bioscience Research Integration Software Platform (*BRISP*) research group is a collaboration between Johns Hopkins School of Medicine, Fraunhofer Center for Experimental Software Engineering, UTHealth- Houston School of Biomedical Informatics, and Biofortis Inc., whose mandate has been to join in the development of a software platform that can help biomedical labs improve their research productivity and satisfaction. Detailed contributions of the four collaborative sites can be seen in Figure 1. Selection of the test lab was based on several criteria including their responsiveness, motivation of the lab's PI, and the richness of lab environment in terms of its ability to represent the manifold changes of use of information technology. A biomedical research information management system called "Labmatrix" developed by Biofortis was used during the project to resolve data management issues in the test lab. Recent studies show that Labmatrix can improve data management capabilities of basic research labs by promoting consistent lab practices, enhancing data integrity and data standardization, allowing collaborative data access, and reducing time spent on specimen and data retrieval by employing query-based retrieval mechanisms [22,23].

The major objective of BRISP collaboration is to use Labmatrix (from now on 'LM') as a tool to improve and systematize productivity of test labs. One important study area reported in this paper include our collaborative efforts to improve data management of high content

screening (HCS) related activities in the test lab. HCS technology revolutionized biological research by enabling nuanced analysis of cell parameters (e.g. cell movement, shape, texture) [24]. However, data generated by HCS usually amounts to multiple terabytes and presents several informatics issues ranging from information store, retrieval, and image mining for pattern identification. Solving these computational challenges required considerable financial resources and thoroughly-constructed IT infrastructure, which are not often available for small and medium sized bioscience laboratories. Data management and integration associated with HCS technology in smaller laboratory settings are an even greater challenge as access to software programmers and systems integrators is often limited. As part of this study, our BRISP group attempted to understand adoption and implementation of Labmatrix to address problems faced by test lab in management of HCS data. Examples of such problems include lack of integrated database with heterogeneous HCS datatypes (numerical data, image data, plate layout), investigator-driven data analysis due to multiple nonintegrated analytics, and problems with interpretation and visualization of results. Complete details about BRISP HCS project is beyond the scope of this paper and can be found elsewhere [25]. In order to resolve some of these issues, BRISP group used several methods to understand the problems, identify possible intervention areas, and monitor implementation progress. The next section presents a brief overview of the methods used in this study over a course of four years.

3. Methodology

Several data collection and analyses methods were employed throughout the course of this collaboration. Important data collection methods included ethnographic observations, semi-structured questionnaires, web-based surveys, face-to-face interviews, conference call summaries, and progress reports that were shared among the project team via Google Docs. The data collection activities were conducted in three phases as described below.

Phase 1- Before the implementation of Labmatrix

Preliminary *ethnographic observations* were conducted to understand the workflow of the bioscience labs, and gain insight into interaction strategies among lab members. The goal was to guide and improve efficiency of data collection in the next phase. A trained researcher unobtrusively observed the activities at different times in the test labs and took observational notes [26]. 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 principal investigators (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 PI of the selected test lab. Both the questionnaires included open-ended and closed specific questions. The questionnaires served as a means to gain knowledge about the overall state of the labs in terms of (1) magnitude and nature of data handled, and (2) data management techniques, and to create an account of current data handling and communication practices in the test lab. 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 [27, 28]. Semi-structured interviews provided an opportunity to learn more about the lab goals and practices. These interviews allowed us to collect detailed descriptions to understand the reasons behind the problems faced by current day biomedical 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 lab data storage, lab 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.

Phase 2- During implementation of Labmatrix

Weekly conference calls and semi-annual face-to-face meetings were conducted to monitor progress of the collaboration in reaching its intended objective. These weekly calls were recorded and transcribed to understand the implementation process of Labmatrix in the test lab. The Conference call summaries were utilized to record, document, and track the progress of the study. We conducted thematic analysis on these summaries to understand the impact of BRISP collaboration on the project outcomes. Thematic analysis is the most common form of analysis in qualitative research [29]. It focuses on examination of patterns or themes within data, where categories emerge from the data [30]. Qualitative coding was used to understand the user experiences of Labmatrix implementation. A total of 261 discussion threads over 52 weeks of conference call summaries were analyzed in this phase.

Phase 3- After implementation of Labmatrix

Web-based surveys and field (laboratory) observations were the data collection activities performed in this phase. The survey questions were designed based on the conference call summaries obtained in the previous phase. Ethnographic observations were utilized to understand the ways in which lab personnel used LabMatrix to support their day-to-day data management activities in the test lab. Given the diversity of our data collection methods, we employed multifaceted techniques grounded in socio-cognitive research to analyze and identify themes as the collaboration advanced with implementation of LM. As newer themes emerged, we used Connection, Communication, Consolidation, Collaboration Interoperability Framework (C⁴IF) [31] to evaluate the interoperability standards supported by information management systems with and without LM. The C⁴IF framework was originally proposed for business information systems as a classification typology. This framework to analyze the interoperability standards of the existing information management practices in a typical bioscience laboratory at granular levels lying underneath the system. The data collected and analyzed using these methods yielded certain interesting and worthy results, as presented below.

4. Results and Discussions

Given the use of diverse analytic methods, we were able to analyze our data from various dimensions related to technical aspects of bioscience lab and research information

management systems, socio-technical factors, and the characteristics of BRISP collaboration during the implementation process.

4.1. Phase 1- Before implementation of Labmatrix

4.1.1 Evaluation of traditional biomedical laboratory research information management systems—

In this phase, we understood research information management profile of the test lab to identify the problems with traditional information systems (such as paper files, excel files, and homegrown databases) based on the practices in the test lab. Based on data from observations, questionnaires, and interviews conducted prior to implementation of LM, we created the information management profile of the test lab. Figure 2 provides a bird's eye view of important research activities usually conducted in the test lab. As seen in the figure, problem areas in data management prior to LM (marked in red flags) were also identified. Identification of challenging areas that existed in the test lab in regards to data management allowed us to narrow down on areas to systematize, and improve using LM in subsequent years. Lack of context in communication, failure to adequately support collaboration, and low interoperability standards were identified as some major problems with traditional research information management applications used by test labs. Limitations to cross study comparisons and data integration were the other challenges faced by biomedical researchers when dealing with spreadsheets, homegrown databases, and so forth with the current information management methods. In addition, we found that poor data organization can potentially lead to substantial data loss as well as degraded security and privacy levels. Myneni et al. provide a detailed analysis of these findings based on the analysis of observational, interview, and questionnaire data [32].

4.1.2. Evolution of Key Performance Indicators—Having identified the problematic areas in the test lab, the BRISP group initially compiled a list of key performance indicators which can help them map their progress as they moved ahead with implementation of LM to address some of these problems. However, these indicators were hardly used as improvement markers as the implementation process progressed and the group realized that this list of indicators would not encapsulate the progress in the test lab. Given the complex nature of research environment in biomedical labs, new performance indicators emerged as the BRISP group progressed with LM implementation in the test lab. *System interoperability and visualization quality* emerged as major indicators of process improvement. For example, Figure 3 depicts how the emphasis laid on visualization increased as BRISP collaboration progressed with implementation of LM in the test lab based on data from conference call summaries.

In addition to visualization quality, the increasing trend of multiple interoperating machines to tackle data management, integration, and interpretation in biomedical labs made system interoperability a vital component for successful deployment of LM in this environment.

4.2. Phase 2- During implementation of Labmatrix

4.2.1 Analysis of Industry-Academia Collaboration and Labmatrix Implementation Process—

The initial face-to-face interviews and questionnaires provided the group with possible intervention and development areas in the test lab.

However, it was not an easy job to pinpoint and focus on one research area. It was only through challenging trial and error that the BRISP group found the high-impact and end-user beneficial project, which is HCS data management. Prior to addressing HCS data management, significant time and effort was applied to other data management issues which were later abandoned. For example, a substantial effort to integrate automated family pedigree-based recording and analysis of hereditary disease data was canceled when the research laboratory realized that the cost in disruption of its current methods would not be sufficiently compensated by a commensurate benefit in research output. Similarly, an effort to automate inventory of antibody and other research samples was not effectively utilized by laboratory personnel because of the perception that the new system required too much time and did not integrate well into the laboratory workflow in ways that benefited both individual laboratory members and the laboratory as a whole. Analysis of the teleconference call content revealed that the lab members became frustrated with the new system during the projects where they did not perceive direct benefit from LM. Given below are sample quotations drawn from conference call summaries illustrating occasional low morale of BRISP collaboration during this course.

“..while he remains enthusiastic about the project, enthusiasm in his lab for doing the BRISP project is low, because typically postdocs, grad students, and other staff have pressing needs and don't see the long term view of the lab..”

“.....we must be talking not about the theoretical but the actual....and these next few weeks are critical, where the amount of time put in per person needs to go up in proportion to whatever is needed to get to this goal, or frankly, the project will be a failure in terms of helping the lab achieve greater scientific productivity and satisfaction.”

Figure 4 provides an overview of the representative tasks performed by the collaboration as part of the study. Of those 261 threads, around 64% of the discussion themes were related to- 1) programming issues and, 2) communication between academia and industry to resolve these issues. Data indicated that this industry-academia collaboration followed a formative iterative implementation and development approach. Regular working group meetings held between system creators and lab members helped the group resolve most issues with the new system. Based on content from the teleconference calls, it was clear that industry personnel offered guidance at various points during the implementation. In turn, regular feedback on the system's performance also helped the industry personnel to augment the system to better respond to the needs of the test lab.

In addition to the assessment of BRISP industry-academia collaboration, we also used a questionnaire tool to identify both positive and negative aspects of LM, in addition to general feedback on system performance. It is important to note that this tool was used during the initial phases of HCS data management project. Results based on this questionnaire are discussed in Section 5.3.

4.3. Phase 3- After implementation of Labmatrix

4.3.1 Assessment of system interoperability standards supported by biomedical research information management practices—We assessed system

interoperability standards with and without LM so as to understand the impact of LM introduction on interoperability standards in the test lab environment. Key findings from interoperability evaluation are presented below. Analysis of the information management practices in the test lab (with and without LM) using C⁴I framework gave us an insight into the interoperability standards of these two different approaches. According to the C⁴I framework, interoperability can be addressed at different levels such as connection, communication, consolidation, and collaboration using this framework. Advancement in each of these areas can influence but cannot determine the advancement of the other areas. A biomedical information management system can have a high degree of interoperability at the communication and consolidation level, while being low at the other two levels. For example, using advanced technologies such as wireless broadband network to exchange data instead of using manual techniques (e.g. compact disks) can be deemed essentially as an advancement in the connection area and this cannot automatically assure semantically rich terminology at the communication/consolidation level and vice versa. Interoperability problems in the context of bioscience lab research lifecycle at (a) connection level indicates issues with data transfer, (b) communication/consolidation indicates problems with data conversion and integration and ambiguous representation of context, (c) collaboration indicates problem with data transfer outside test lab environment, context-sensitive agility of data conversion for easy communication with collaborators. An analysis of the degree of interoperability supported by the data management practices with and without LM is included in Table 1. Comparative ratings were given on a three point scale of Low – Medium- High scale based on our analysis of data from observations and interviews.

Data sharing options were limited in the test lab, however, the lab members were using some means (such as email, common server/driver storage) to transfer their research data to their collaborators within the lab as well as at the distributed research sites. As reflected in Table 1, the connection level of this system had a medium (M) level of interoperability. But data ontology and format were not considerably standardized, thus giving a low (L) interoperability in the consolidation and communication domain. Similarly, collaboration was minimally supported by the system, although significant was achieved “manually” usually through formal meetings, electronic mail and shared documents rather than through integrated information management systems. For an information management system to succeed in collaborative interoperability, common ideas on workflow patterns and functions were needed to be established [5], which were not realized in this case. Hence, it was given a low (L) rating for interoperability at the collaboration level. Myneni et al. provided detailed analysis of interoperability levels supported by traditional information systems in the test lab before LM implementation [33].

After the introduction of LM, we again conducted observations and interviews in the test lab. Since LM is a web-based system, access control was used to create accounts for lab members and other collaborators such that data can be accessed and exchanged appropriately. Hence, this layer was given a high rating. Since lab members had to perform manual data pre-conditioning steps such that they can upload the information from other software to LM, and this process is still somewhat tedious, consolidation-communication was given a medium ranking. LM was used with other interoperating applications to provide data visualization at an improved quality level in the test lab. This showed that LM

integrated well with other applications for adequate and comprehensive data analysis. However, there was still a scope for improvement in terms of programming agility and ease of establishing a collaboration layer between systems. Therefore, the collaboration was ranked medium. Data showed that the interoperability of research environment provided by LM is higher than other traditional research information management systems.

Although LM provided better support for system interoperability, the BRISP group often encountered problems with data format, data upload, and data loss during migration from the old system to LM. Past work on this project revealed that the test lab members were sometimes frustrated during the implementation process of LM [32]. In order to understand the reasons behind lab member's frustration, we analyzed the implementation process to understand user experience and industry-academia collaboration in this study.

4.3.2 Evaluation of user perspectives on Labmatrix—Four LM users responded to the web-based questionnaire, where three participants responded to all the questions. Two of the four lab members have rudimentary experience of two years with LM, while the others have less exposure to the new system. As part of the questionnaire, the users were asked to rate LM in different categories (self-explanatory, easy to learn, market standards, fun to work with, similarities with other contemporary systems, and aesthetic look) on a four point scale of Poor-Excellent (see Figure 5). In addition to rating LM with respect to the above-mentioned qualities, the users were also asked to give a brief account of the advantages and limitations associated with LM. The users mentioned the new system as a one-stop resource for their research data management and analysis with strong querying capabilities. However, results indicated poor maneuverability, difficult navigation, and non-intuitive user interface often made working with LM unpleasant.

In addition, we conducted direct observations in the test lab to understand the nature of limitations with LM. We observed the lab members as they carried out their regular tasks using LM. During these observations, we also focused on how LM was being integrated into lab's workflow. These data provided us with valuable insight into two new concepts called “Human interoperability” and “Cognitive interoperability”

4.3.3. Evaluation of human component of information system interoperability and its effect on implementation process—With the introduction of LM into HCS workflow, lab members were able to accomplish certain tasks which they were never capable of earlier. Examples of such capabilities include as 1) looking at all their data sets together, 2) integrating and analyzing different types of data, and 3) visualizing the data in several different ways. Above all, the query module of LM presents a new paradigm of allowing people with limited database programming skills to ask complex questions against complex datasets. Detailed explanation of this particular HCS project can be found here [25] and is beyond the scope of this paper. The new workflow added additional data preconditioning steps to the old process, and the lab members had to learn to use the new system in order to conduct their daily tasks using LM. Learning such new skills and acquiring knowledge pertaining to the redefined workflow could be attributed to user frustrations during implementation process. In addition, the lab members had to work with multiple systems (such as excel files, laboratory server, Labmatrix) and work with each one

of these applications simultaneously to populate custom data forms in LM. Analyses of interview data showed that LM users were very concerned with highly complex nature of the new process.

The frustrations caused by the learning curve associated with a system may be alleviated by incorporating design principles that support human interoperability, which promotes the inclusion of certain system characteristics during system design that allow human beings to use other similar tools with minimal training, showing some generalizable skills [34]. Human interoperability may be highly intertwined with another concept called cognitive interoperability, which is defined as the human being's way of thinking when using a system [33]. A perfect alignment between the human being's expectations of a system and system response would likely reduce lab members' resistance to new technology.

One lab member perceived LM as a “Query Results Analyzer”- a tool to analyze all results of a particular query, and hence wanted to have all the results related to a “search” query be highlighted and displayed as main content of the screen for easy analysis. However, the system was designed in such a way that the query area was emphasized rather than the query results. The designer viewed the system as “Query Machine”. This indicated a clear discrepancy between the thought patterns of the laboratory member and system designers. This lack of “Cognitive Interoperability” might be a reason behind user frustration with LM.

In summary, scientific research was greatly affected by the traditional data management practices prevailing in the test lab. Traditional information management systems resulted in sub-optimal collaborative environment, inadequate support for comprehensive data analysis, and provided lower interoperable standards. The BRISP group attempted to intervene with LM in order to systematize data management in test lab. As the group progressed with the LM implementation, they tracked the project progress by establishing performance indicators. Data revealed that LM provided the test lab with a better interoperable environment, higher data integration, visualization, and querying capabilities. It is important to note that additional workflow steps, failure to foresee added value of LM by lab personnel, unexpected programming problems affected the implementation process. In a nutshell, BRISP group faced challenges during this implementation, but managed to create a better research environment in the test lab by providing standardized data management and novel analysis mechanisms.

5. Summary of Findings and Discussions

In this paper, we present the efforts made by a multi-site distributed industry-academia collaboration to systematize and improve research information management in a biomedical academic test laboratory. Our evaluation of BRISP collaboration has revealed multiple findings that bring together multiple informatics challenges, ranging from socio-technical factors, research collaboration, personnel management, information retrieval and management technicalities associated with bioscience labs.

Phase 1- Before implementation of Labmatrix

The traditional research data management systems were inadequate in providing support for dynamic environment in biomedical laboratories that deal with voluminous, high content, heterogeneous data, and new ways of data management became necessary. The major problems with the traditional systems were related to data maintenance, data exchange within and outside the laboratory, interdisciplinary collaboration, and publications. Personnel-driven information management activities posed several problems to record-keeping and overall research productivity of the laboratory. The ways in which suboptimal data management can affect different operations of a bioscience lab were understood. Although BRISP project is oriented towards understanding specific challenges faced by a single test lab, the results are consistent with findings from other studies [16,17]. Industry-driven standards, ontology models to manage metadata, and platforms that enable integration of heterogeneous data sources can resolve certain issues identified in our study. This is one of the To our knowledge, this is one of the first reported studies that investigated information management and its relation to lab outcomes in terms of productivity and satisfaction of scientific investigators, as suggested in recent publications [35,36]

Phase 2- During implementation of Labmatrix

The BRISP collaboration faced problems because of unanticipated programming obstacles, time delays, user dissatisfaction, and lack of demonstrability of technology success. Such problems lengthened implementation time and lowered morale of BRISP group. Important cornerstones of the BRISP collaboration which helped the team advance with the implementation of Labmatrix in the test laboratory without many hurdles include motivating leadership, involved personnel from academia and industry, and regular monitoring of project status. While reported studies on similar collaborative efforts in bioscience environment have focused on providing detailed technical architectures of system development, our study focused on aspects of human learning and incremental nature of BRISP collaboration. The BRISP group identified certain performance indicators to track their overall progress throughout the implementation process. The performance indicators in a biomedical laboratory environment are complex and emerging in nature. Therefore, key indicators were revised and adjusted as the implementation process evolved. In our case, system interoperability and visualization emerged as important markers of improvement. However, user perceptions of Labmatrix based on these two indicators were not measured explicitly since they do not encapsulate the overall relevance of Labmatrix to test lab projects in which the members are directly involved, and the direct benefits of these indicators are not immediate.

Phase 3- After implementation of Labmatrix

Labmatrix was able to assist the test lab to integrate, analyze, and visualize their research data. Data integration, strong querying procedures with comprehensive data management were some key capabilities that test laboratory was able to achieve with Labmatrix implementation. New features were added to Labmatrix in order to customize it to suit the needs of the test lab members in an iterative fashion. Examples of such new capabilities included visualization, optimized data normalization routines, and strong query module that

present a totally new paradigm allowing laboratory members with minimal database programming skills to build complex queries on complex biomedical datasets. Labmatrix provided a better interoperable research environment in the test lab. However, there is further scope for Labmatrix to improve its capability to support interoperability at communication level of C⁴I framework by developing common ontology and establishing well-accepted data schema. In addition, although Labmatrix has systematized data management in test lab, user perceptions of direct benefits of Labmatrix indicate there is still a room for improvement. Lack of demonstrability of benefits of informatics solutions such as Labmatrix is always a challenging task and this resulted in abandoning of initial projects by multiple lab members, subsequently may have affected their perceived level of benefits. Minimal level of perceivable direct benefit for the laboratory and/or for individual's work was a major factor overall project success. Lessons learned from this collaboration are similar to recent reports on the implementation studies of electronic health records [37, 38]. Bioscience information management systems are similar to any other health information technology, where it is essential to establish both long-term and short-term measurement variables that are socio-technical and project-driven in nature. The nature of barriers we have identified with information management procedures (prior to Labmatrix implementation) in the test lab are (a) cognitive (e.g. lack of streamlined procedures, labor intensive, lack of visualization), (b) social (e.g. data sharing issues, communication ambiguities), (c) organizational (e.g. no direct personal benefit, financial constraints, personnel management), and (d) systemic (e.g. use of homegrown databases, lack of industry-driven/discipline-specific standards). Implementation of Labmatrix has addressed multiple barriers through data visualization, web-based technology-driven lab workflow, centralized shareable data warehouse, and customized lab-specific standards across personnel. Use of formative usability evaluation methods [39] can enable integration of user and technology usage context into the system, improving overall technology acceptance. Methods like heuristic evaluation and cognitive walkthrough [39-41] have been proven to be beneficial in improving system usability. While BRISP collaboration has predominantly focused on understanding test lab needs, prioritizing feasible and highly desirable intervention areas, developing test lab-specific algorithms and query capabilities, similar collaborative efforts should focus on integration of usability techniques with development process so that one can mitigate user frustrations with counterintuitive system features and misalignment of computerized protocols with lab workflow.

6. Conclusions and Lessons Learned

We present below some important lessons learned during Labmatrix implementation that can help any academic biomedical laboratory migrating to a new information system plan ahead of their implementation process. In addition, these pointers can also provide guidance to academia-industry collaborations in biomedical research in terms of expectations and possibilities.

1. The acceptability of any new system in a biomedical laboratory depends on user understanding of the system. In the world of technology dominance, human beings still play an important role for successful task accomplishments. Evaluating the human dimension of implementation process is essential for optimizing the

advantages of any new technology. Such assessment can help us improve system acceptability and user satisfaction.

2. Most times, the system being implemented in a biomedical laboratory needs to be customized to better suit the needs of the end users. Such customization results in augmented system features such as a) improved interface design with better usability, and b) new capabilities including visualization, and c) optimized programming algorithms. Complex data schemas, relational data models, and intuitive visualization of query results are often seen as grand challenges to achieve data integration in biomedical arena [43, 44]. Industry-academia collaborative efforts such as this can provide rich, mutually-beneficial venues to researchers and system developers investigating these problems.
3. Given the functional complexities to be supported by information management systems in biomedical laboratories, no single system will be successfully able to solve all data management tasks at hand. In the future, multiple technologies working together will form the basis of information systems' operational framework. Therefore, interoperability should be realized as one important built-in capability of any future information management tools.
4. Working closely with system creators, having regular feedback loops, and developing clear “as-is” and “should-be” depictions of laboratory scenarios at the start of the implementation process can help resolve most issues that arose during this collaboration.
5. High system interoperability with other software, quick iterative technology development, workflow-driven implementation, industry-academia interaction, dedicated personnel are some essential qualities any collaboration should have to optimize new technology implementation process.
6. In addition to machine interoperability, human and cognitive interoperability plays a major role in system acceptance and can improve user experience of an implementation process. The better the cognitive interoperability match of humans with systems, the more would be the system acceptance. Similarly, higher human interoperability can help end users transfer their skills to use new technology, thus mitigating their learning curve with the new system.
7. End users need to take time to explore the new system completely in order for them to appreciate its capabilities. An end-user will be satisfied with a system if it is customized to his/her needs, however; such customization will induce additional programming tasks for the industry to work on. It is essential to realize that developing optimized programming algorithms takes time and an end user should be patient enough and continue using the new system to accomplish their regular work during this time lapse.
8. While in an industry-academia collaboration, it is important to realize that priorities of both parties might sometimes be different. Industry may limit itself from too much customization of the system, given the need to address priorities across its client base.

9. While novelty of a software application can have a positive impact on user perception, adoption, and satisfaction, lack of workflow-centered evaluation of such applications might impede their acceptance. Usability evaluation of these software suites is essential to eliminate cognitive, social, systemic, and organizational barriers.
10. It is important for a biomedical laboratory to carefully define what their needs are before starting the technology migration/transfer process. In the BRISP collaboration, the test labs did not really know what they needed or wanted, and that definition of the need evolved through the process, which is probably not the ideal way to proceed, but probably is not uncommon. Part of the issue might have also been due to the fact that the lab came into this collaboration as a “test lab,” not as a client who was going out to look for a solution to an already well-defined problem.
11. Proper understanding of end users' needs is essential for success of any informatics project. However, novel ventures such as BRISP collaboration might have difficulty understanding these needs and can result in initial failures. It is important to realize that there is an intrinsic learning curve associated with these failures. Learning from failures and using those failures to build newer knowledge resources is vital. Such knowledge provides valuable guidance to the next generation biomedical informatics researchers.

Acknowledgments

The study was supported in part by a grant (1R41CA105217-01A1-STTR) from National Institute of Health/ National Cancer Institute (NIH/NCI). Support was also provided by generous gifts from the Guerrieri Family Foundation and from Mr. and Mrs. Clarice Smith. We thank all study participants for their valuable time and contributions.

References

1. ConsortiumIHGS. The human genome. *Nature*. 2001;860–921.
2. Sansone SA, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, Hide W. Toward interoperable bioscience data. *Nature genetics*. 2012; 44(2):121–126. [PubMed: 22281772]
3. Louie B, Mork P, Sanchez FM, Halevy A, Hornoch PT. Data integration and genomic medicine. *Journal of Biomedical Informatics*. 2007; 40:5–16. [PubMed: 16574494]
4. Dennis C. Biology databases: Information overload [Electronic Version]. *Nature*. 2002 May.14:417.
5. Anderson NR, Lee ES, Brockenbrough JS, Minie ME, Fuller S, Brinkley J, et al. Issues in biomedical research data management and analysis: Needs and barriers. *Journal of American Medical Informatics Association*. 2007; 14:478–488.
6. Zhang, Z.; Townsend, JP.; Yu, J.; Cheung, KH.; Bajic, VB. Data integration in bioinformatics: current efforts and challenges. INTECH Open Access Publisher; 2011.
7. Anderson NR, Ash JS, Hornoch PT. A qualitative study of the implementation of a bioinformatics tool in a biological research laboratory. *International Journal of Medical Informatics*. 2007; 76:821–828. [PubMed: 17084664]
8. Sujansky W. Heterogeneous database integration in biomedicine. *Journal of Biomedical Informatics*. 2001; 34:285–298. [PubMed: 11977810]
9. Tenenbaum JD, Sansone SA, Haendel M. A sea of standards for omics data: sink or swim? *Journal of the American Medical Informatics Association*. 2014; 21(2):200–203. [PubMed: 24076747]

10. Ding L, Wendl MC, McMichael JF, Raphael BJ. Expanding the computational toolbox for mining cancer genomes. *Nature Reviews Genetics*. 2014; 15(8):556–570.
11. Baralis E, Fiori A. Exploring heterogeneous biological data sources. Paper presented at the 19th international workshop on database and expert systems application, 2008 DEXA'08 Turin. 2008 Sep.
12. Maghrabi F, Faheem HM, Soliman T, Fayed ZT. A multiagent-based framework for integrating biological data. *International Journal of Intelligent Information Technologies*. 2008; 4(2):24–36.
13. Viksna J, Celms E, Opmanis M, Podnieks K, Rucevskis P, Zarins A, et al. PASSIM - An open source software system for managing information in biomedical studies. *BMC Bioinformatics*. 2007; 8
14. Milsted AJ, Hale JR, Frey JG, Neylon C. LabTrove: A lightweight, web based, laboratory “blog” as a route towards a marked up record of work in a bioscience research laboratory. 2013
15. Wang X, Liu L, Fackenthal J, Cummings S, Olopade OI, Hope K, et al. Translational integrity and continuity: Personalized biomedical data integration. *Journal of Biomedical Informatics*. 2009; 42(1):100–112. [PubMed: 18760382]
16. Miller PL, Nadkarni PM, Kidd KK, Cheung K, Ward DC, Banks A, et al. Internet-based support for bioscience research: a collaborative genome center for human chromosome 12. *Journal of the American Medical Informatics Association*. 1995; 2(6):351. [PubMed: 8581551]
17. Anderson NR, Ash JS, Tarczy-Hornoch P. A qualitative study of the implementation of a bioinformatics tool in a biological research laboratory. *International journal of medical informatics*. 2007; 76(11):821–828. [PubMed: 17084664]
18. Li, H.; Gennari, JH.; Brinkley, JF. AMIA Annual Symposium Proceedings. Vol. 2006. American Medical Informatics Association; 2006. Model driven laboratory information management systems; p. 484
19. Jakobovits, R.; Soderland, SG.; Taira, RK.; Brinkley, JF. Proceedings of the AMIA Symposium. American Medical Informatics Association; 2000. Requirements of a Web-based experiment management system; p. 374
20. Fong, C.; Brinkley, JF. AMIA Annual Symposium Proceedings. Vol. 2006. American Medical Informatics Association; 2006. Customizable Electronic Laboratory Online (CELO): A web-based data management system builder for biomedical research laboratories; p. 922
21. Afsarmanesh H, Belleman RG, Belloum ASZ, Benabdelkader A, van den Brand JFJ, Eijkel GB, et al. VLAM-G: A Grid-based virtual laboratory. *Scientific Programming*. 2002; 10(2):173–181.
22. Suh, KS.; Remache, YK.; Patel, JS.; Chen, SH.; Shaikh, AM.; Wang, J., et al. Paper presented at the American medical informatics association translational bioinformatics summit. San Francisco, CA: 2008. Translational bioinformatics-guided workflow for procurement of patient samples.
23. Gopalan, B.; Orloff, M.; Stanuch, K.; Waite, K.; Platzer, P.; Eng, C. Paper presented at the American medical informatics association translational bioinformatics summit. San Francisco, CA: 2008 Mar. Integration of clinical, molecular and cellular phenotypes to unravel the mechanisms of complex diseases.
24. Comley J. High content screening. *Drug Discovery*. 2005; 31
25. Berlinicke CA, Ackermann CF, Chen SH, Schulze C, Shafranovich Y, Myneni S, et al. High-Content Screening Data Management for Drug Discovery in a Small-to Medium-Size Laboratory Results of a Collaborative Pilot Study Focused on User Expectations as Indicators of Effectiveness. *Journal of laboratory automation*. 2012; 17(4):255–265. [PubMed: 22357564]
26. Van Maanen, J. Ethnography. In: Kuper, A.; Kuper, J., editors. *The social science encyclopedia*. 2nd. London: Routledge; 1996. p. 263-265.
27. Bernard, R. *Research methods in anthropology: Qualitative and quantitative methods*. Walnut Creek: AltaMira Press; 2002.
28. Crabtree, B.; Miller, W. *Doing qualitative research*. Newbury Park, CA: Sage; 1992.
29. Boyatzis, RE. *Transforming qualitative information: Thematic analysis and code development*. Sage; 1998.
30. Patel VL, Kaufman DR, Allen VG, Shortliffe EH, Cimino JJ, Greenes RA. Toward a framework for computer-mediated collaborative design in medical informatics. *Methods of Information in Medicine*. 1999(38):158–176. [PubMed: 10522119]

31. Peristeras V, Tarabanis K. The connection, communication, consolidation, collaboration interoperability framework (C⁴IF) for information systems interoperability. *Interoperability in Business Information Systems*. 2006(1):61–72.
32. Myneni S, Patel VL. Organization of Biomedical Data for Collaborative Scientific Research: A Research Information Management System. *International Journal of Information Management*. 2010; 30:256–64. [PubMed: 20543892]
33. Myneni, S.; Patel, VL. Proceedings of the American Medical Informatics Association Annual Symposium (AMIA). San Francisco, California: 2009. Assessment of Collaboration and Interoperability in an Information Management System to Support Bioscience Research; p. 463–467.
34. Goldkuhl, G. The challenges of Interoperability in E-government: Towards a conceptual refinement. *Pre-ICIS SIG eGovernment Workshop*; Paris: 2008.
35. Burgun A, Bodenreider O. Accessing and Integrating Data and Knowledge for Biomedical Research. *Yearb Med Inform*. 2008;91–101. [PubMed: 18660883]
36. Mirel B, Eichinger F, Nair V, Kretzler M. Integrating Automated Workflows, Human Intelligence and Collaboration. *Summit Translat Bioinforma*. 2009; 2009:79–83.
37. Ash JS, Anderson NR, Tarczy-Hornoch P. People and organizational issues in research systems implementation. *Journal of the American Medical Informatics Association*. 2008; 15(3):283–289. [PubMed: 18308986]
38. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. *BMC health services research*. 2010; 10(1):231. [PubMed: 20691097]
39. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. *Computers in biology and medicine*. 2002; 32(3):141–149. [PubMed: 11922931]
40. Nielsen, J. Conference companion on Human factors in computing systems. *ACM*; 1994 Apr. Usability inspection methods; p. 413–414.
41. Preece, J.; Rogers, Y.; Sharp, H. *Interaction design: Beyond human–computer interaction*. New York: Wiley; 2002.
42. Jeffries, R.; Miller, JR.; Wharton, C.; Uyeda, K. Proceedings of the SIGCHI conference on Human factors in computing systems. *ACM*; 1991 Apr. User interface evaluation in the real world: a comparison of four techniques; p. 119–124.
43. Nagarajan, K.; Ahmed, RM.; Phatak, A. Proceedings of the 30th VLDB conference. Vol. 2004. Toronto, Canada: 2004. Database challenges in the integration of biomedical data sets; p. 1202–1213.
44. Li A. Facing the Challenges of Data Integration in Biosciences. *Engineering Letters*. 2006; 13(3): 5–8.

Figure 1. Four major research sites in "Bioscience Research Integration Software Platform" Industry- Academia collaboration

Figure 2. Profile of data management in the academic biomedical test laboratory

Timeline for Labmatrix implementation in Test Laboratory

Figure 3. Evolution of “visualization quality” as a key performance indicator during Labmatrix implementation in test laboratory over 12 month period

Figure 4. Classification of discussion themes based on the content of conference calls over a period of 12 months

Figure 5. Ratings given by members of the test laboratory for different qualities of Labmatrix

Table 1
Analysis of system interoperability supported by the information management practices
in the test laboratory II (with and without Labmatrix) using C⁴I framework

Layer name in the C ⁴ I framework	Objective of the layer	Interoperability rating before Labmatrix implementation	Interoperability rating after Labmatrix implementation
<i>Collaboration</i>	Support for action towards a shared goal	Low	Medium
<i>Consolidation/Communication</i>	Commonly accepted entity relationships or data schema, data format, data meaning	Low	Medium
<i>Connection</i>	Data exchange via compact discs, internet, and so on	Medium	High