

# Operationalizing Personalized Medicine: Data Translation Practices in Bioinformatics Laboratories

Maryam Ficociello

*Beedie School of Business, Simon Fraser University, Burnaby, BC, Canada., ficociello@sfu.ca*

Ellen Balka

*School of Communication, Simon Fraser University, Burnaby, BC, Canada., ellenb@sfu.ca*

Follow this and additional works at: <http://aisel.aisnet.org/amcis2012>

---

## Recommended Citation

Ficociello, Maryam and Balka, Ellen, "Operationalizing Personalized Medicine: Data Translation Practices in Bioinformatics Laboratories" (2012). *AMCIS 2012 Proceedings*. 17.

<http://aisel.aisnet.org/amcis2012/proceedings/PerspectivesIS/17>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2012 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Operationalizing Personalized Medicine: Data Translation Practices in Bioinformatics Laboratories

**Maryam Ali Ficociello**  
School of Communication  
Simon Fraser University

8888 University Drive, Burnaby BC, V5A 1S6  
Canada  
[ficociello@sfu.ca](mailto:ficociello@sfu.ca)

**Ellen Balka**  
School of Communication  
Simon Fraser University

8888 University Drive, Burnaby BC, V5A 1S6  
Canada  
[ellenb@sfu.ca](mailto:ellenb@sfu.ca)

## ABSTRACT

This paper presents findings from an ethnographic study of two genomics and bioinformatics labs. The focus of this research is on the day-to-day practices of using multiple technologies to integrate data across different platforms. We argue that sociotechnical challenges (including technical, contextual, and political challenges) emerge when data integration practices are carried out, due to the embedded nature of the important, yet unrecorded and implicit historical information that each dataset carries. We observed that sociotechnical sensemaking was common place in lab work, and was the only method for working out the complexity of the challenges which arose during data integration activities. We suggest that due attention be given to this matter, as challenges related to assessing data are likely to arise once more when such data travels back to the bedside, where it is poised to directly impact human health.

## KEYWORDS

genomics, bioinformatics, personalized medicine, translational research, sense-making, sociotechnical practices, science, technology and society studies.

## INTRODUCTION

A promising area in the practice of medicine is that of personalized medicine (e.g. Bottinger, 2007; Cascorbi, 2010; Evans, Meslin, Marteau, and Caulfield 2011; Lesko, 2007), defined as “the application of genomic and molecular data to better target the delivery of health care, facilitate the discovery and clinical testing of new products, and help determine a person’s predisposition to a particular disease or condition” (Abrahams, Ginsburg, and Sliver, 2005: 396). In personalized medicine, knowledge of a person’s genetic and molecular profile has the potential to guide preventive care, and the selection of therapies that are most effective and safe (Abraham et al. 2005). This has major implications for revolutionizing the health care system, through cost reductions (Fackler and McGuire, 2009) and improved health outcomes. Research in the area of personalized medicine is data intensive. The creation of knowledge from data requires researchers to integrate large and diverse data sets, which presents numerous challenges in areas such as data representation, and the linking of heterogeneous data sets (data integration) (Louie, Mork, Martin-Sanchez, Halevy, Tarczy-Hornoch, 2007). Indeed, data integration has been considered a major deterrent to the success of personalized medicine (Louie et al. 2007; Payne, Embi, and Chandan, 2009). Despite the promise of data integration platforms, adoption has been low, owing in part to sociotechnical barriers, data ownership and data security (Payne et al. 2009). As efforts to wed genetic data to electronic medical records are reported in the popular press (MacArthur, 2011), scientists are quick to point out that few examples of the utility of personalized medicine exist in clinical settings (Laurence, 2009).

Few studies to date have considered the sociotechnical issues arising from the practices of data integration in its most raw form: at the bench side (i.e. in health science laboratories). Here, we report findings from research focused on the day-to-day practices of using new technologies in genomics and bioinformatics labs to carry out scientific research which, we suggest, will be essential to the translation of lab based scientific discoveries to health system use. We argue that without consideration of the underlying processes involved in the data integration work at the micro level, the often-promised benefits of personalized medicine will remain elusive.

With backgrounds in information system research, and science, technology and society studies, we developed a research project aimed at exploring practices associated with the use of genomics data within the context of personalized medicine. This approach provided a unique vantage point from which to study sociotechnical challenges associated with data integration. Although translational bioinformatics has received only limited attention amongst scholars and practitioners in information systems research, this domain offers a unique opportunity through which to examine emergent challenges associated with the increasingly prevalent practices of creating software tools and data resources intended for use in multiple contexts. Our findings provide insight into the increasingly common practices of integrating multiple data sources, in order to gain insight about a variety of issues. We have identified three levels of sociotechnical challenges that emerge when using multiple tools and technologies in scientific labs, where local data integration practices are common place. Material reported here seeks to fill gaps in knowledge about sociotechnical challenges which arise with data integration, by reporting about the many data integration and sense-making practices we observed in our study sites.

Below, we provide an overview of the theoretical fields which frame our approach. This is followed by an overview of our methodology. We then present our findings and analysis, followed by a discussion of our results. We end with a conclusion and recommendation for further research.

## BACKGROUND

Our current work- which is concerned with data integration practices of pre-clinical genomic scientists<sup>1</sup> can be understood in relation to three related literatures: translational bioinformatics, information infrastructures, and the sociotechnical approach. Each is outlined briefly below.

### Translational Bioinformatics

Achieving the goals of personalized medicine relies upon the movement of genomics and bioinformatics research from bench (science labs) to bedside (clinical settings). This ‘movement’ is commonly referred to as translational research or translational bioinformatics. The term translational research refers to the need to move research from ‘bench-to-bedside’ by improving the interface between basic science and clinical medicine (Woolf, 2008). Translational bioinformatics refers to “the development of storage, analytic, and interpretive methods to optimize the transformation of increasingly voluminous biomedical data, and genomic data in particular, into proactive, predictive, preventive, and participatory health.” (AMIA, n.d.). Translational bioinformatics sits between and (ideally) interoperates with electronic medical records, clinical research informatics systems, statistical analysis and data mining. It includes—and draws on—the areas of genomic sciences, biomedical sciences, and health informatics.

The majority of literature about translational bioinformatics has addressed macro level aspects of this translation, including a range of topics in the ethical, legal, and political domains related to the need to develop new policy and governance structures in the genomics field (Gaskell and Bauer 2006, Atkinson, Glasner, and Lock 2009). These are not insignificant issues. However, while important, they are not the focus of our current inquiry. Payne et al. (2009: 131) suggested that although the need to collect, manage, integrate, analyze and disseminate large-scale heterogeneous (biomedical) data sets is a common theme in translational bioinformatics research, “well-established and broadly adopted theoretical and practical frameworks and models intended to address such needs are conspicuously absent.” The study we report about here was undertaken with this in mind, and focuses on the issues and challenges arising in relation to the management and integration of large scale heterogeneous biomedical data.

### Information Infrastructures

Definitions of information infrastructures (also referred to as cyberinfrastructures (CIs) and eScience) abound. They vary in the extent to which infrastructural issues are conceptualized as technical problems to be solved or as sociotechnical processes. In information systems (IS) research, early definitions of information infrastructures were provided by Hughes (1983) who defined them as large technical systems with sociotechnical power structures. For the purposes of this study, we define information infrastructures as “the set of organizational practices, technical infrastructure and social norms that collectively provide for the smooth operation of scientific work at a distance” (Edwards et al., 2007:6). This definition recognizes the

---

<sup>1</sup> It is noteworthy to mention here that our two labs are ‘research labs’, which often experiment with new and emerging tools, many of which are new to researchers’ practice. Research labs such as those we studied differ from diagnostics labs and bedside data use. In contrast to research labs where new tools and procedures are constantly being used and procedures may not be repeated over and over, in diagnostic labs the same procedures are carried out repeatedly so the scientists are very well versed with the data handling and analysis they are performing and the software they use. Presently there are few implementations of genomic data use in bedside settings, so it remains unclear to what extent clinicians working at the bedside are cognizant of data issues which may have occurred ‘upstream’ from use.

inter-relatedness of the social and technical in the production of information infrastructures, and also recognizes the contingent nature of technology. Star and Ruhleder (1996) characterized information infrastructures as emerging with five properties: they are embedded, transparent, have a reach or scope, can be learned as part of membership in communities, link with conventions of practice, allow for the embodiment of standards, are built on an installed base, and become visible upon breakdown. More recently in information systems research, the term digital infrastructure is used (Tilson, Lyytinen and Sorenson 2010).

### Sociotechnical Approach

Our basic object of study in this research is not a single element, but rather a network of heterogeneous elements, including technology, people, and institutions. Considering a sociotechnical approach allows us to identify the many elements at play in translating bioinformatics research. The sociotechnical perspective has been widely researched in the information systems field (Bostrom and Heinen 1977, Kling and Saachi 1982, Mumford 1983, Orlikowski 1992, Lin and Cornford 1995, Avgerou 2002, Orlikowski and Scott 2008). Here we define the sociotechnical approach as the interrelatedness of ‘social’ and ‘technical’ aspects of an organization or a society as a whole (Trist and Bamforth 1952). The term ‘technical’ does not necessarily refer to material artifacts, but is used in the broader sense and incorporates technicalities, including structures, formalities and procedures. Two main premises of this approach are that organizations need to keep in mind that we cannot expect the ‘social’ to behave like the technical’ (i.e. people are not machines), or vice versa (machines, even if inscribed with human-like procedures, cannot have a mind of their own). The second principal is that we need to strive for a joint-optimization of both the ‘social’ and ‘technical’ aspects of a given system, in order to realize their full potential (Walker et al. 2004).

## METHODOLOGY

### Ethnographic Case Studies of Two Genomics and Bioinformatics Labs

In an effort to meet our commitment to our funder<sup>2</sup> to produce deliverables of strategic importance to the regional genomics research community, we conducted research in two genomics labs in order to gain micro-level insights about the day-to-day work practices of scientists. Our data collection begun in September 2011 and will continue until August 2012. We used several data collection methods, including semi-structured interviews with researchers, *in situ* observations of researchers at work, analysis of documents pertaining to regular meetings (lab presentations and lab meeting minutes), and data from two widely-used discussion forums about bioinformatics, to gain insights about genomic scientists’ everyday work practices. These will be supplemented by interviews with genomic scientists and stakeholders external to the labs, about issues arising in working with genomic data, including ethical and legal issues. Here we are reporting only on our ethnographic work in the two lab settings; which we refer to here as the ‘Alpha’ and ‘Zeta’ labs.

Ethnographic, participant observations were carried out in two laboratory settings, during regular weekly and bi-weekly lab meetings, and were supplemented by day-to-day observations of researchers working in both the wet and dry genomics and bioinformatics labs. For purposes of data collection, we identified the unit of analysis as data handling processes surrounding two specific technologies that are commonly used in both lab settings: Perl programming language and Structured Query Language (SQL) in the dry lab. Our focus during observations was on how these technologies were used to handle and analyze data, how data were used, produced, and integrated within or across different technological platforms, and identifying the sociotechnical challenges that arose with these practices.

Semi-structured interviews are underway with 28 lab members. Research participants are a combination of graduate students, postdoctoral fellows, lab technicians, bioinformaticians (web and database developers), and tenured faculty. Field notes were taken during observations, and later transcribed. All interviews are recorded, and transcribed as well. Data analysis was undertaken on an ongoing basis. All data has been formally coded (Strauss and Corbin 1990, Miles and Huberman 1994) using qualitative data analysis software NVivo 9. All research participants gave their informed consent to take part in this research.

In the next two sections, we present the objects of particular interest to our discussion here: the two lab case studies, and the technologies that were at the centre of most of the scientists’ research.

---

<sup>2</sup> We were funded through Genome BC’s Strategic Opportunities Fund #4.

### *Description of Alpha and Zeta Labs*

Although the Alpha and Zeta Labs are situated in the same department and institution, each lab is slightly different in terms of research orientation and focus. Alpha Lab consists of an interdisciplinary wet lab and dry lab (computer lab) environment used to investigate pathogenomics questions. Computational analyses of genomics are combined with individual proteins and other lab data to facilitate experimenting with new hypotheses and testing them in model host systems<sup>3</sup>. At the time of this research, the lab was comprised of 12 members.. The staff members were primarily responsible for the ongoing maintenance of either the wet or dry lab. There was a wet lab technician, as well as web/database developers involved in managing the four databases that were created and continue to be supported by the lab. The remaining 8 lab members were mostly scientists-in-training in the form of graduate students and postdoctoral fellows. The majority of research that was observed in this lab was carried out in the dry computer lab, as only 1 postdoctoral fellow occasionally carries out work in the wet lab. Within the Alpha Lab, our observations were comprised of two settings: general lab meeting observations and lab work observations. The ethnographer (MAF) attended 18 lab meetings, and conducted 5 dry lab work observations. Each lab meeting lasted an hour on average, while lab work observations lasted an average of 2 hours. The ethnographer's role as participant observer included volunteering to help with testing the usability and functionality of two of the lab's databases.

In comparison, the Zeta Lab's objective is to develop bioinformatics programs and tools for understanding genomic architecture and expression. There are two parts to this lab, a dry lab and a wet lab both of which are regularly used. In the wet lab, *Caenorhabditis elegans* (*C. elegans*) worms were used to carry out experiments. Dry labs are where computers (supported by powerful servers and grid technology) were used to run genomic computations. At the time of this research, the lab consisted of 17 members, including volunteers, master's students, PhD students, postdoctoral fellows and lab technicians. A total of 19 general lab meeting observations, 7 wet lab work observations, and 3 dry lab work observations were carried out. Lab meeting observations lasted an average of 3 hours, while the wet and dry lab observations lasted 2.5 hours on average.

In both labs, during the lab meetings, members would report about recent research progress to the lab director, who mentored and supervised students and staff, led discussions, and was quite involved in helping with on-going analyses of data, fine-tuning of scientists' methods, and carrying-out quality-control checks of the scientist's research standards. The lab meetings allowed each member to share their best practices in relation to tool use. Constructive feedback was constantly given to researchers along with advice about possible future research avenues they might pursue. In contrast to this setting, the wet and dry lab work settings were quite different. Here, the scientists mostly worked in solitude on their individual projects. Collaboration with other scientists was commonplace, but mostly in the form of trouble shooting in the dry lab, or carrying out a part of an experiment in the wet lab.

### **Technology in Use: Perl and SQL**

Through our ethnographic work, we came to understand that most of the work of integrating data from multiple sources in our two lab field sites was accomplished with the use of two technologies: Perl programming language and SQL database quarrying language. Both of these technologies are situated in the dry labs. The Perl programming language is a high-level, general purpose, dynamic programming language that has become widely used by biologists in bioinformatics research. It was originally developed in 1987 by Larry Wall as a general purpose scripting language to make report processing easier (Sheppard 2000). Perl facilitates easy manipulation of text files, with its ability to processes and detect patterns in data. The release of Perl version 5, which supports object-oriented programming, made it much easier to develop reusable modules of biology across research centers (Tisdall 2000). Perl was commonly used in the dry lab for string processing of biological data like gene or protein sequences. The core work of dry lab scientists was in the manipulation of textual data sets from multiple databases and resources. Structured Quarry Language (SQL), on the other hand was vital for running quarries across databases. Both programming languages allowed scientists to access, extract, divide, or insert data sets from one or a combination of databases and convert them to other forms of output. Many scientists in bioinformatics did not necessarily receive formal training in the use of Perl or SQL, yet these two languages (amongst others) had become such important tools in bioinformatics in recent years that scientists routinely learned how to use them either on their own or from peers.

### **FINDINGS AND ANALYSIS**

Here we discuss findings which emerged from observation of day-to-day practices primarily in the dry lab. We identify three levels of sociotechnical challenges that lab scientists face in their day-to-day work practices involving technology. These

---

<sup>3</sup> The Alpha lab used both human and non-human data, and the sociotechnical challenges mentioned here applied to both types of data.

sociotechnical challenges include: *technical challenges*, *contextual challenges*, and *political challenges*, each of which is more complicated than the previous type. These challenges are outlined below, using examples from our fieldwork.

### Technical Challenges

Technical challenges are those that can be resolved with the enhancement or redistribution of resources. Examples include attracting suitable personnel, training and developing researchers, providing better access to information, investing in high performance computing infrastructure at the local level, or purchasing software.

An example of a technical challenge is that **databases were often ‘down’ and/or out of service**. This meant that scientists could not carry out their usual work until the databases or sites were functional again. This type of problem was caused by either maintenance work on those sites, or servers being out of order due to poor resource support. This was mostly observed with a database named *Geoffrey*. Another technical issue was related to the **knowledge and training level** of researchers, especially in relation to computer science proficiency. As most of these scientists were biologists or chemists by training, they either had a minor degree (or a few courses) in computer science, or learned bioinformatics techniques as they practiced their research. It was hence not surprising to witness repeated examples of scientists facing obstacles related to bugs in a piece of code they wrote.

These challenges were further aggravated by the **fast evolving nature of the bioinformatics field**, which presented new tools and methods often and demanded that scientists learn how to use them. Obstacles also emanated from the **poor design** (e.g., broken links, missing complete protein or gene curations) and **usability functions** (e.g., difficulty in navigating a given database interface) of many databases and/or tools. Researchers often identified bugs in different databases and sites, and reported them back to developers. Additionally, **poor documentation of tool development or best practices** concerning usability meant that knowledge of usability issues could only be obtained through practice, or learned informally through ‘socializing’ with other users. Scientists routinely turned to verbal conversations for trouble shooting, and used online discussion forums as a means of finding solutions to common problems. Although this informal sharing may on the surface appear to be a waste of time, in an environment where documentation may be absent or out of date, verbal communication and use of discussion groups are what allow scientists to overcome problems which would otherwise obstruct their work (Orr 1992).

When researching dry lab practices, some reoccurring obstacles were so common that they became taken-for-granted by the researchers. An example of this was evident when a scientist was trying to figure out why her SQL query was not yielding the results she expected. She described that this ‘starring at the screen’ (in attempting to make sense of the obstacle) is what she does most of her day. Although taken for granted, these technical challenges were not minor: during our observations they had a direct effect on delaying the continuity of participant’s work, and led to frustration. The scientists we observed explained that many of the technical challenges resulted from the fact that tools and databases were developed for short-term projects. Limited time and funding resources are invested in making tools bug and error free, or more user friendly or functional. This created multiple problems when the same tools were subsequently used by masses of users, often at a distance from where they were developed. These sorts of technical challenges were influenced by the broader political context, which (for example) came to bear on funding, and hence influenced documentation, tool development, etc.

### Contextual Challenges

Contextual issues presented a second level of challenge. These stem from the combination of more than one technical challenge, or the emergence of unforeseen contextual issues that may arise with the use of databases. Unlike technical challenges that can be resolved with the increase in resources, in order to resolve contextual challenges, there is a need to coordinate and cooperate with other stakeholders in the community through knowledge sharing, or collaborative practices.

Many of the contextual challenges were faced when integrating data from **multiple sources**, such as from different databases. Another contextual challenge was the **knowledge inscribed into the technology**. For example, knowing that a tool was developed for a given purpose makes clear what limitations and biases that tools might have if used in a different context for a different purpose. The **quality** of output of different tools also differed substantially. There was often mention of how certain sequence alignment tools provided different (and sometimes biased) results. There is no gold-standard for which tool is best used for which type of sequences, which had multiple consequences for the scientists. We observed one scientist comparing data from one sequencing technology with an isolate that had been sequenced with a different sequencing technology. It appeared that the scientist had to be *mindful* of issues such as which alignment tools were used to generate particular sequences, which meant that two sequences could naturally differ. Hence, this would not necessarily mean that any missing segments would be representative of mutations or deletions in a given sequence read, but may rather reflected the use

of different tools in sequencing. The researchers were ‘trained’ to consider such biases of tools when running analyses. Knowledge of the context in which tools were developed was essential to making sense of genomic data.

Another contextual challenge arose in relation to **integrating data** and the lack of data standards. Because of the diverse types of databases and tools that are constantly developed by different scientists, it is often critical to integrate data from multiple sources in order for them to achieve meaningful results. This means that complicated quarries need to be written specifically to make such integrations of multiple data formats from multiple databases. This usually also entails layers of data cleaning activity, and other adjustments that need to be carried out in order to have results that are truly useful. **Collaborative practices** are one of the leading reasons that bioinformatics is developing at such a quick pace. However, these same collaborative practices do result in contextual challenges. While scientists share their data and tools with others in hope of **recognition for their contribution**, those very tools often reflect limitations (e.g., of resources) at the local level, which result in very **little standardization and/or stabilization**. This relates to the third type of challenge we observed, which is that of the political will to fund projects over a long period of time so that the stabilization efforts are possible in the form of reconfiguring a generic system to local practices (Bjiker and Law, 1992). This would allow for the continuity of work that would result in more sustainable efforts at the local level that could be used (and hence translated) at the inter-organizational domain. We elaborate on this challenge next.

### Political Challenges

The third level of challenges and issues that arise at the local level of lab work in bioinformatics can broadly be categorized as political challenges. These challenges encompass the wider environmental subsystem that influences the bioinformatics domain. It includes the rules and regulations, whether formal or informal, that govern relationships between lab scientists, their respective disciplines, their collaborators, and society at large. In order for these challenges to be resolved, change in regulation or policy has to occur either within or outside of the organizations involved. A common characteristic of these political challenges is that they tend to have long-term implications, and require a long time to resolve.

Funding and data ownership were two political challenges of greatest concern to scientists. Scientists in our study were concerned that funders preferred to go for the ‘low-hanging fruit’. That is, to fund projects that would guarantee results in the short term. This meant that many projects with a longer life span would not be funded simply because their completion was incompatible with funders’ needs to demonstrate results. Also, projects that were more exploratory in nature were difficult to fund, since the results were not as certain. This phenomenon led to the clustering of ‘over researched’ topics, while other areas of bioinformatics were left under-explored. Another funding-related problem was that of **discontinuity**. Once projects ended, it was difficult to obtain funding for ongoing maintenance or further development of existent databases or other tools that were devised for the project. Hence, it was impossible to maintain the benefits of a project’s results after it ended. This resulted in many databases of limited usefulness being available to the public. This was closely related to the concern that funders did not seem to realize the **importance of bioinformaticians** as key contributors to the maintenance and sustainable advancement of research tools, which related to the ‘articulation work’ (the invisible work required to keep things running smoothly, see Strauss 1985) they perform.

**Data ownership** was also an issue of concern. There was careful recognition given to data used from other researchers, as credits were often acknowledged both verbally as well as in publications. Latour and Woolgar (1979) have pointed to this phenomenon using the term ‘cycles of credits.’ There was also some concern about data produced as part of projects that were funded by pharmaceutical companies. This concern was related to the ownership being given to the private sector funders, which meant that results were locked for their own use. Scientists were concerned that while they would not be able to carry out research without such funding, that reliance on funding from companies could delay the potential benefits of the research if the data generated through the research was not made available beyond the funder. Such practices could possibly delay wider drug development or even the potential eradication of disease. Closely related to data ownership were questions of ethical considerations, which are of considerable interest and importance but beyond the scope of this paper.

### DISCUSSION

As many technical challenges can be resolved with an increase in resources, of greater concern to this study is addressing the second level of challenges outlined above: the *contextual challenges*. Our findings suggest that this particular level of challenges is especially sociotechnical in nature: they consist of both social and technical elements, which interact in complex ways (Bygstad 2008). In a more general sense, the *contextual challenges* can be classified as a data integration problem (Louie et al. 2007), which entail critical collaborative challenges as well. What makes practices of data integration challenging is the need for tacit or implicit knowledge at the local level of practice, which need to be re-created or instantiated for collective organizational learning, as well as for work to carry on and reach a completion stage. The knowledge we refer to is characterized as ‘not immediately tellable’ (Orr 1996: 119). This problem has been cited before,

where researchers turned to focusing on tacit knowledge and reflective practice (Schön 1991; Lanzara 1991; see also Yanow and Tsoukas 2005); perhaps as a solution to the problem of ‘tell-ability’ (Yanow 2006:1747). The term ‘articulation work’ has also been coined to explain this phenomenon, defined as “a kind of supra-type of work in any division of labor, done by the various actors” Strauss (1985: 8). Strauss (1985) describes articulation work as containing three parts: work involving the meshing of numerous tasks, work involving meshing the tasks of multiple unit workers (individuals, departments, institutions etc.), and the meshing of actors with their various types of work and implicated tasks (Strauss 1985).

We have observed that there is much articulation work that is carried out due to the distributed nature of sharing data and technologies across the bioinformatics community. There are of course many practices that can be objectified and recorded and made visible, and then passed on along with the data and technologies, and which could yield predictable results if applied. These are in the form of protocols, formal structures, plans, and procedures or schemas (Schmidt and Bannon 1992). However, the real concern of our research is the phenomenon observed by Gerson and Star (1986) as local and temporary practices, which within themselves require articulation work. Gerson and Star (1986: 266) point out that “reconciling incommensurate assumptions and procedures in the absence of enforceable standards is the essence of articulation. Articulation consists of all the tasks involved in assembling, scheduling, monitoring, and coordinating all of the steps necessary to complete a production task. This means carrying through a course of action despite local contingencies, unanticipated glitches, incommensurate opinions and beliefs, or inadequate knowledge of local circumstances.”

New technology (such as those created commonly in the bioinformatics community) results from the coordination of vast networks of people, machines, processes, and materials, what Latour (2005) calls *actor networks*. The technologies we live with seldom represent the only technological solutions possible; but rather, represent the outcome of a complex series of negotiations, which occurred within a web of both human and non-human actors (e.g., the installed base of a computer system). Technology development and use are both social processes, whose outcomes depend upon social and technical factors, yet the need for and consequences of these social factors is often under-recognized in scientific and technological development processes. Lab scientists, though perhaps lacking concepts such as articulation work, are very aware of the consequences if it is not performed. Yet the invisibility of that work arguably contributes to difficulties in getting it funded.

In response to the need for articulation work, it seemed like scientists engaged in a new form of articulation work, that of sociotechnical sensemaking. We use the term sociotechnical sensemaking to refer to those practices that scientists engage in when trying to find solutions to obstacles that have no known or documented solution. We denote this sensemaking as specifically ‘sociotechnical’, since there was a constant consideration of reasons for why a given obstacle might have arisen, both on a social as well as a technical level, or both levels at once. Scientists considered for example, if a given problem was caused by inherent database structures, or by purely social causes such as the practice followed by colleagues in gene annotation practices. They used sociotechnical mechanisms to make sense of the problems they faced and to overcome them. Examples include troubleshooting by reviewing technical specifications related to the database/tool in use, or more socially by asking more experienced colleagues or turning to web-based discussion forums for answers. We refer to these underlying processes of trying to make sense of the issue at hand, as sociotechnical sensemaking. It echoes the work of Payne and Mendonka (2007) who argue for the importance of attending to the semiotic nature of the day-to-day processes of translating research information to practice and practice information to research.

Issues and challenges related to data re-use and data integration have been identified in several areas including electronic health records (Tolar & Balka, 2012) and enterprise resource planning systems (Volkoff, Strong and Elmes 2005; Light and Wagner 2007). In their discussion of secondary usage of EMR data, Tolar and Balka identified successful information infrastructure implementation and the ability to sustain initiatives associated with data use as pre-conditions for secondary use of data in a primary care setting. Work undertaken during our current study suggests that similar issues arise in pre-clinical lab settings, where data collected have demonstrated that several areas identified by Tolar and Balka (the need for software functionality and knowledge of software, the need to test and tailor data to fit the new purpose, (data cultivation), and the need for designated personnel and other resources to sustain activities related to secondary use of data) are required to sustain data integration activities in the pre-clinical labs which are our focus here. Work reported here, together with previous work concerned with issues arising during data integration and data re-use activities suggests that although the specific activities required to support data integration activities may vary from setting to setting, it may nevertheless be useful to conceptualize these varied activities as ‘sociotechnical sensemaking’ processes, and to recognize that time, attention and resources must be allocated to support these activities during certain kinds of new technology implementations.



## CONCLUSION

This study essentially reports on the dilemma that has emerged with the development of technology, and the rise of the Internet. It has become a common practice for scientists to share their data, databases and other locally created tools and technologies across organizational boundaries, and in fact such practices are often mandated when development has been supported with public-sector funds. However, these very practices of sharing data and technologies present problems at the receiving end. Shared tools are only able to carry limited amounts of information that allow extra-local users to make sense and contextualize the data that they produce and carry. This is because data and databases, when shared, are abstracted from the context of their development, which can lead to a host of problems (see Schuurman and Balka, 2009 for a more in depth discussion of this phenomenon). We refer to these challenges as *contextual challenges*.

In order to understand how these *contextual challenges* are being resolved or worked out, we conducted ethnographic research to identify and understand the mechanisms undertaken on a day-to-day basis by scientists to make sense of, and eventually overcome such challenges. The issues we have identified may be magnified as data move via information infrastructures away from the laboratory environment (bench-side) and into clinical settings (bed-side). The greater the distance a data user is from the supports required for sociotechnical sensemaking, the higher the probability that an end user of data (e.g., a clinician interpreting test results) will have access to information that might allow the end user to make sense of particularly unusual or possibly irregular findings in data. In the absence of recognition that articulation work is occurring and is essential to the success of translational bioinformatics, adequate resourcing of such activities is unlikely, and development of methods for capturing and making explicit the tacit knowledge and practices (articulation work) required to use emerging and rapidly changing information infrastructures, is likely to remain inadequate. Yet these issues are likely to grow in importance as the distance from the lab to the bedside increases. It is very likely that healthcare professionals at the bedside will lack an understanding of the affordances and constraints of the tools used to develop data, or the biases inherent to data produced with particular sequencing tools, which may have contributed to tests or diagnostics they will be expected to use. Hence the failure to recognize the need to support articulation work practices in labs not only poses a threat to research (and possibly data quality) in lab settings, but also threatens the quality of medicine to be delivered “at the bedside” as front line practitioners will lack access to the very information (e.g., documentation about limitations inherent to a particular dataset which was used to formulate a particular diagnosis) they will require to assess the integrity of results they obtain in clinical use of tools.

This research has implications for the development of tools that can translate research findings cultivated at the bench-side, for their use and application at the bed-side, through personalized medicine. Without the consideration of these often ignored contextual challenges, the goal of operationalizing personalized medicine may be undermined by inappropriate interpretations of data, and are likely to take longer to achieve.

## ACKNOWLEDGMENTS

We would like to thank the researchers from the Alpha and Zeta labs who have generously given their time up for us to understand their work processes. We are also grateful to Genome BC for their funding support through the Strategic Opportunities Fund #4 competition, and the Assessment of Technology in Context Design Lab, which hosted this project.

## REFERENCES

1. Abrahams, E., Ginsburg, G. S., Silver, M. (2005) The personalized medicine coalition: Goals and strategies. *American Journal of Pharmacogenomics*, 5, 345–355.
2. AMIA (No Date). *AMIA Strategic Plan*. Retrieved March 22 2011 from <https://www.amia.org/inside/stratplan>.
3. Atkinson, P., Glasner, P. and Lock, M. (2009) *Handbook of genetics and society: Mapping the new genomic era*, Routledge, London.
4. Avgerou, C. (2002) New socio-technical perspectives of IS innovation in organizations, in C. Avgerou and R.L. LaRovere (Eds.): *ICT innovation: Economic and Organizational Perspectives*, Edward Elgar, Cheltenham.
5. Bijker, W. and Law, J. (1992) *Shaping technology building society: Studies in Sociotechnical change*. The MIT Press, Cambridge, MA, USA.
6. Bostrom, R.P., and Heinen, J. S. (1977) MIS problems and failures: A socio-technical perspective, Part 1: The causes, *MIS Quarterly*, 1, 3, 17-32.

7. Bottinger, E. P. (2007) Foundations, promises and uncertainties of personalized medicine, *Mount Sinai Journal of Medicine*, 74, 1, 15-21.
8. Bowker, G. and Star, S. L. (1999) *Sorting things out: Classification systems and its consequences*, MIT Press, Cambridge, MA, USA.
9. Bradley, E. H., Curry L. A., and Devers, K. J. (2007) Qualitative data analysis for health services research: Developing taxonomy, themes, and theory, *Health Services Research*, 42, 1758-1772.
10. Cascorbi, I. (2010) The promises of personalized medicine. *European Journal of Clinical Pharmacology*, 66, 749–754.
11. Edwards, P., Jackson S., Bowker, G., and Knobel, C. (2007) Understanding cyberinfrastructure: Dynamics, tensions and design: Report of a workshop on “History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures” Ann Arbor, MI: University of Michigan.
12. Evans, J. P., Meslin, E. M., Marteau, T. M., Caulfield, T. (2011) Deflating the genomic bubble. *Science*, 331, 861-862.
13. Fackler, J. L., McGuire, A. L. (2009) Paving the way to personalized genomic medicine: steps to successful implementation. *Current Pharmacogenomics and Personalized Medicine*, 7, 2, 125- 132.
14. Gaskell, G., and Bauer, M. W. (2006) *Genomics and society: legal, ethical, and social dimensions*, Earthscan, London.
15. Gerson, E. M., and Star, L. S. (1986) Analyzing due process in the workplace. *ACM Transactions on Office Information Systems*, 4, 3, 257-270.
16. Ginsburg, G.S., Willard, H. S. (2009) Genomic and personalized medicine: foundations and applications. *Translational Research*, 154, 6, 277-287.
17. Hughes, T. P. (1983) *Networks of power: Electrification in Western Society 1880-1930*, John Hopkins University Press.
18. Kirrily, R. (2012) Perlintro: a brief introduction and overview of Perl. Retrieved 2012-02-17 from <http://perldoc.perl.org/perlintro.html>.
19. Kling, R. and Scacchi, L. (1982) The web of computing: Computer technology as social organization, *Advances in Computers*, 21, 1-90.
20. Lanzara, G. F. (1991) Shifting stories: Learning from a reflective experiment in a design process, in D. A. Schön (Ed.) *The reflective turn: Case studies in and on educational practice*, Teachers College Press, New York, USA, 285–320.
21. Latour, B. (2005) *Reassembling the social: an introduction to actor-network-theory*, Oxford University Press, Oxford, UK.
22. Latour, B. and Woolgar, S. (1987) *Laboratory life: The construction of scientific facts*, Princeton University Press. Princeton, USA.
23. Laurence, J. (2009) Getting personal: The promises and pitfalls of personalized medicine, *Translational Research*, 154, 6, 269-271.
24. Leonelli, S. (2010) Machine science: The human side, *Science*, 330-317.
25. Lesko, L. J. (2007) Personalized medicine: Elusive dream or imminent reality?, *Clinical Pharmacology & Therapeutics*, 81, 6, 807-816.
26. Light, B., and Wagner, E. (2006). Integration in ERP environments: Rhetoric, realities and organizational possibilities, *New Technology Work and Employment*. 21(3), 215-228.
27. Louie, B., Mork, P., Martin-Sanchez, F., Halevy, A., Tarczy-Hornoch, P. (2007) Methodological review: Data integration and genomic medicine. *Journal of Biomedical Informatics*, 40, 1, 5-16.
28. MacArthur, D. (2011) When ‘cautious’ means ‘useless’. *Wired*, Retrieved 2011-02-18 from <http://www.wired.com/wiredscience/2011/02/when-cautious-means-useless/>
29. Miles, M. B. and Huberman, A. M. (1994). *Qualitative data analysis*, Thousand Oaks, Sage, CA, USA.
30. Mumford, E. (1983) *Designing human systems*, Manchester Business School, Manchester, UK.
31. Orlikowski, W. (1992) The duality of technology: Rethinking the concept of technology in organizations, *Organization Science*, 3, 3, 398-427.
32. Orlikowski, W., and Scott, S.W. (2008) The entangling of technology and work in organizations, in *Working Paper #168, London School of Economics and Political Science*.
33. Orr, J. (1996) *Talking about machines: An ethnography of a modern job*. ILS Press.

34. Payne, P. R. O., Embi, P. J., Chandan, K. S. (2009) Translational informatics: Enabling high-throughput research paradigms, *Physiological Genomics*, 39, 131-140.
35. Payne, P. R. O., Mendonca E. A. (2007). Conceptual knowledge acquisition in biomedicine: A methodological review, *Journal of Biomedical Informatics*, 40, 582-602.
36. Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H., Arnheim, N. (1985) Enzymatic amplification of beta-glob in genomic sequences and restriction site analysis for diagnosis of sickle cell anemia, *Science*, 230, 4732, 1350–1354.
37. Schmidt, K., and Bannon, L. (1992). Taking CSCW seriously: Supporting articulation work, *Computer Supported Cooperative Work (CSCW): An International Journal*, 1, 1, 7-40.
38. Schuurman, N. and Balka, E. (2009). Alt.metadata.health: Ontological context for data use and integration. *Journal of Computer Supported Cooperative Work*, 18, 1, 83-108.
39. Sheppard, D. (2000). Beginner's Introduction to perl, Retrieved 2012-02-17 from <http://www.perl.com/pub/2000/10/begperl1.html>
40. Star, S. L., and Rudheler, K. (1996). Steps towards an ecology of infrastructure, *Information Systems Research*, 7, 1, 111-134.
41. Strauss, A. L. (1985) Work and the division of labor, *The Sociological Quarterly*, 26, 11985, 1-19.
42. Strauss, A. L. and Corbin, J. (1990) Basics of qualitative research, Sage Publications, Inc; 3rd edition.
43. Schön, Donald A. (ed.) (1991) The reflective turn: Case studies in and on educational practice. New York: Teachers College Press.
44. Tilson, D, Lyytinen K, and Sorenson, C. (2010). Digital Infrastructures: The missing IS research agenda. *Information Systems Research*. Article in advance, P. 1-12.
45. Tisdall, J. (2000) Mastering BioPerl for bioinformatics, Retrieved 2012-02-17 from <http://oreilly.com/catalog/mpperlbio/chapter/ch09.pdf>
46. Tolar, M. and Balka, E. (2012). Caring for individual patients and beyond: Enhancing care through secondary use of data in a general practice setting. *Int. J. Med. Inform.* Article in advance.
47. Trist, E.L. and Bamford, K.W. (1951) Social and psychological consequences of the Longwall method of coal-getting, *Human Relations*, 4, 3-28.
48. Volkoff, O., Strong, D. M., & Elmes, M. B. 2005. Understanding enterprise systems-enabled integration. *European Journal of Information Systems*, 14 (2), 110-120.
49. Walker, B.H, Holling, C.S., Carpenter, S.R., and Kinzig, A. (2004). Resilience, adaptability and transformability in social–ecological systems, *Ecology and Society*, 9, 2, 5.
50. Woolf, S. H. (2008) The meaning of translational research and why it matters, *JAMA*, 299, 2, 211-213.
51. Yanow, D. (2006) Talking about practice: On Julian Orr's talking about machines, *Organization Studies*, 27, 12, 1743–1756.
52. Yanow, D., and Tsoukas, H. (2005) Reflecting in/on practice, Presented at the EGOS (European Group on Organizational Studies) annual conference, Berlin, 30 June–2 July 2005.