CLINICAL AND TRANSLATIONAL SCIENCE INFORMATICS INFRASTRUCTURE

A Framework and Case Study

Arkalgud Ramaprasad, Annette L. Valenta
University of Illinois at Chicago, Chicago, IL 60607, USA

Ian Brooks
National Center for Supercomputing Applications, Urbana, IL 61801, USA

Keywords: Clinical and Translational Science, Informatics Infrastructure.

Abstract: This paper presents a comprehensive socio-technical framework for the design and development of a Clinical and Translational Science Informatics Infrastructure (CTSII). Based on our experience with developing and applying the framework we present a case study to illustrate the issues that arise in the creating a CTSII, and how possibly these issues can be resolved. The framework is presented as a menu with six columns, each column representing a dimension of the framework. The categories within each dimension can be concatenated, with the conjunctive phrases/words between the columns, to form sentences that describe all the functions of the CTSII. Elucidation of all the combinations will provide an exhaustive list of all the possible functions of CTSII.

1 INTRODUCTION

In 2002, the National Institutes of Health (NIH) in the US charted a roadmap for this century to identify opportunities and gaps in biomedical research in order to make the biggest impact on the progress of medical research (NIH Office of Communications, 2003). The roadmap seeks to foster new pathways to discovery, to develop innovative research teams of the future, and to reengineer the clinical research enterprise (NIH Office of Portfolio Analysis and Strategic Initiatives, 2006). It seeks to create a new discipline called Clinical and Translational Science (CTS) to reduce the time-to-practice of biomedical scientific discoveries, and the time-to-research of clinical and community health care issues (NIH Office of Portfolio Analysis and Strategic Initiatives, 2007).

Clinical and Translational Science (CTS) by definition is interdisciplinary; however, it is difficult to foster interdisciplinary research cutting across basic, animal, clinical, and public health disciplines. One barrier to such research is disciplinary silos that often manifest themselves in the form of departments, colleges, journals, and conferences. A well designed CTS Informatics Infrastructure (CTSII) can help break these barriers.

It is natural for people to know more about the research and researchers in their discipline than in others. Disciplinary research is the foundation of academic advancement, at least in the short run. The incentives systems in universities are woven around disciplinary productivity and the performance is evaluated by peers in the discipline. Consequently, the silos foster relationships within their boundaries rather than across them. While disciplinary research is necessary, it is also necessary to cut across these silos to develop CTS. How can CTSII help?

There is a disconnection between the availability and the use of informatics tools and techniques. Many popular consumer informatics tools demonstrate immense potential. Our objective is to import these tools and techniques and apply them to create an effective CTSII. Metaphorically, the ideal CTSII is a combination of Google™, Facebook™, Amazon™, and Orbitz™. It should have the global indexing, ranking, and search capabilities of Google™; the social networking capabilities of

Ramaprasad A., L. Valenta A. and Brooks I. (2008).
CLINICAL AND TRANSLATIONAL SCIENCE INFORMATICS INFRASTRUCTURE - A Framework and Case Study.
In *Proceedings of the First International Conference on Health Informatics*, pages 213-218

Facebook™; the data mining, cataloging and customer [researcher] relationship management of Amazon™; and the complex scheduling [chaining] capabilities of Orbitz™. Analogues of these four systems, which have revolutionized consumer informatics, will serve as excellent bases for the design of CTSII (Valenta et al., 2007).

2 CTSII FRAMEWORK

To break the silos of research, while simultaneously advancing science, the CTSII should facilitate back and forth translation of information between basic researchers, animal researchers, clinical investigators, and public health researchers. (We use information to generically connote data, information, and knowledge.) It must support the translation of information between the sub-disciplines of each group as well (Valenta et al., 2007).

The quality and quantity of information translation will determine the effectiveness of CTS. In the following sections we present a systematic framework to analyze and design CTSII. We are currently using the framework to develop the CTSII at our university. The framework incorporates, integrates, and extends the ideas from the CTS proposals that have been funded by NIH. It has been presented to and discussed with a large group of researchers across the campus from a wide range of disciplines – including medicine, nursing, applied health sciences, engineering, business administration, public health, and pharmacy. We will discuss the issues that have been raised during our discussions, and how we plan to address them.

CTSII is not just a technological infrastructure, but also a social, psychological, organizational, and educational one – a fact that can be easily overlooked. The proposed system will, by its very design, restructure workgroups, causing stress to the organization, its social groups, and individuals. Appropriate education, consultation, training, change management, evaluation, and assessment mechanisms will be critical for the success of CTSII.

2.1 CTSII Menu

We present our CTSII framework as a menu with six columns (Figure 1), each representing a dimension of the framework. The six dimensions represent: (a) the different types of integration central to translation, (b) the different areas of research that have to be translated, (c) the resources available for translation, (d) the diseases that form the focal point of translational research, (e) the methodological steps in any research (including translation research), and (f) the tools for translation. In fact, the menu is a method of representing a matrix with six dimensions; each dimension being represented by a column. The categories within each dimension can be concatenated, with the conjunctive phrases/words between the columns, to form sentences that describe all the functions of the CTSII. Some example combinations follow:

- Lateral integration of basic research databases related to HIV/AIDS for theory construction using scientist relationship management.
- Temporal integration of public health researchers related to asthma for empirical testing using scientific workflow management.

It can be seen that even with the abbreviated list of entries in the columns, the total number of combinations is very large, indicating the complexity of CTSII. Elucidation of all the combinations will provide an exhaustive list of all the possible functions of CTSII. It would be difficult, if not impossible, to incorporate all of them in one system – they have to be prioritized. The following provides a description of the six dimensions and a sample of the categories within each.

2.1.1 Integration Dimension

Integration is one of the major driving forces behind CTS. It has been a somewhat elusive but important goal sought through earlier initiatives in interdisciplinary and multi-disciplinary research. The objective of CTS is to substitute serendipitous integration with systematic integration.

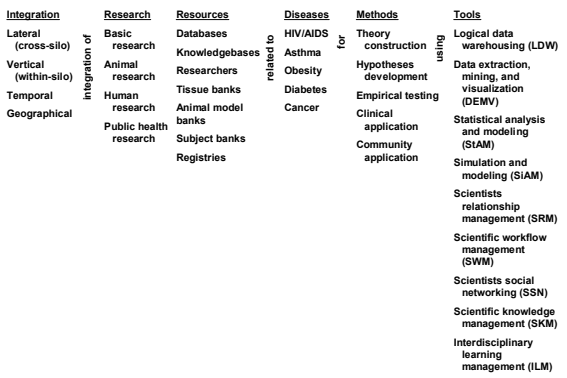

Figure 1: CTSII Menu.

The integration dimension has four categories. They are: (a) Lateral integration (cross-silo); (b) Vertical integration (within-silo); (c) Temporal integration (over time); and (d) Geographical integration (across many locations).

CTS requires integration across and within silos of basic research, animal research, human research, and public health research. To be effective, the research also has to be integrated longitudinally – over time, and across many geographical locations where the research resources may be located. Hence the four categories of the integration dimension.

Each of the four types of integration imposes a different set of requirements on the CTSII. In addition to the purchase and installation of the hardware, software, and networks, the participants will have to be informed, and trained to use the new infrastructure, and the processes of scientific collaboration will have to be reengineered to utilize the new infrastructure.

2.1.2 Research Dimension

CTS encompasses four types of research. They are: (a) Basic research; (b) Animal research; (c) Human research; and (d) Public health research.

Each of these phases includes many components; for example, the human research phase includes human trials, treatment modalities, and clinical practice. Similarly, public health research includes dissemination of the results to the public and community.

These four phases encapsulate the concept of moving basic research to the patient's bedside and the public – the central tenet of CTS. While these four phases are commonly presented as a progressive sequence, research ideas may originate in any phase and move across these phases in any order. Thus, research ideas may originate from basic research and may be fed-forward directly into human research; or, they may originate in public health research and may be fed-back directly to animal research.

One of the major concerns of CTS is that each phase tends to be a silo. These silos are reinforced by norms of the disciplines and associated incentives. The silos inhibit feed-forward and feed-back. Consequently, the movement across the phases tends to be slow and not smooth. A significant body of research may accumulate in a phase without any impact on the subsequent phases through feed-forward or on prior phases through feed-back. When this happens, both the creative and corrective value of feed-forward and feed-back is lost. Streamlining the feed-back and feed-forward mechanisms using CTSII, on the other hand, can improve both the efficiency and effectiveness of translation (Ramaprasad, 1979, 1982, 1983). Similarly, silos within silos can inhibit feed-in. Streamlining feed-in using CTSII can lead to improvement in the quality of feed-forward and feed-back.

2.2 Resources Dimension

CTS requires integration of a large number of resources. They are: (a) Databases – central, homegrown, relational, flat files, etc.; (b) Knowledgebases – structured, unstructured, text, formal, informal, etc.; (c) Researcher databases – directories, résumés, profiles, etc.; (d) Tissue banks; (d) Animal model banks; (e) Subject banks – deidentified subjects, identified subjects, volunteers, etc.; and (f) Registries

Under each of the above categories, there is likely to be a large number of subcategories, and ultimately a larger number of actual resources. Developing an inventory of these resources will be a key step in developing the informatics infrastructure.

Researchers often focus exclusively on the integration of databases as a requirement of the CTSII. While databases are important, integration of information about other resources is equally important. A clinical researcher, for example, probably does not need access to the genomic database used by a basic researcher to discover a gene marker for breast cancer, but needs information about the marker and how to test for it. The clinical and basic researchers need to know which other researchers can help them develop a reliable test for the marker, how they can obtain a panel of subjects for a trial, and the tools to evaluate the results of the trial.

The CTSII should, ideally, replace the usually ad-hoc processes for accessing these resources with more efficient and effective processes. By making the resources visible and accessible to all, the CTSII should improve both the quality and utilization of these resources.

2.3 Diseases Dimension

Different diseases may require different combinations of capabilities in the CTSII. While gene-based diseases may require the ability to handle large volumes of genomic data, environmentally induced diseases may require the ability to manage disparate public health data. Similarly, some diseases may require the ability to

educate and interact with the public health workers and the community physicians and nurses.

The CTS in an institution may be focused on a few or a large number of diseases. Following is an illustrative list of diseases suited to the current research at our university: HIV/AIDS, asthma, obesity, diabetes, and cancer.

Focusing on the above will fit our university's expertise as well as its mission as a premier urban research university. Other institutions may have other priorities based on their vision, mission, strategy, and environment. The design of the CTSII will naturally depend upon the different informatics required for the management of the diseases in question.

2.4 Methods Dimension

The informatics requirements of different stages of research are different. The CTSII should support all the stages.

The first three categories in the methods dimension are standard stages in research methodology. Carrying them forward to clinical and community application, the last two categories, are an essential part of translation. Thus, the five categories are: (a) Theory construction; (b) Hypotheses development; (c) Empirical testing; (d) Clinical application; and (e) Community application.

The type of translation required for translational theory construction may be quite different from that required for translational clinical application. The former may require metaphorical translation from one discipline to another; the latter may require a methodological translation. (Please see section 3 below for a more detailed discussion of the different types of translation.) Thus, for the metaphorical translation, the CTSII may have to facilitate the social networking of theoreticians from the different disciplines, while for methodological translation, the methodologists from the disciplines may have to be brought together. These groups, in turn, may need access to different types of knowledge.

2.5 Tools Dimension

CTS requires integration using many tools. This dimension articulates the metaphor we used earlier for CTSII as a combination of Google™, Facebook™, Amazon™, and Orbitz™. Under each of the following categories, there is likely to be a large number of tools. In fact, many tools may span multiple categories. Developing an inventory of these tools will be a key step in developing the informatics infrastructure. The categories of tools include:

- Logical data warehousing tools;
- Data extraction, mining, and visualization tools;
- Statistical analysis and modeling tools;
- Simulation and modeling tools;
- Scientist relationship management tools;
- Scientific workflow management tools;
- Scientist social networking tools;
- Scientific knowledge management tools; and
- Interdisciplinary learning management tools.

The above is not an exhaustive list of the types of tools. There are many other types of tools, and many more are likely to emerge in the future. There will also be many more tools within each type. The difficulty is not one of the availability of tools, but of their application to CTS and in developing workflow management capabilities to integrate the tools. The cross-fertilization of the application of these tools across traditional CTS disciplines, and between non-CTS disciplines (for example: marketing, production and operations management, semiotics, computer science, and library science) will be facilitated by the CTSII.

3 CTSII CASE STUDY

In this section, we will present the key issues that have arisen as we have tried to adopt and apply the CTSII framework described above over the past year. These issues are unlikely to be unique to our institution. It is intrinsic to the nature of transformation that CTS is trying to engender. In addition to illuminating the process of application, the case also highlights the importance of considering the psychological, social, organizational, and educational aspects of the CTSII.

3.1 Not Just Databases

To many, the term informatics appears to connote only databases. A number of early meetings focused exclusively on developing an inventory of the databases and making them easily accessible to other researchers. Perhaps it reflected the participants' primary concern with their research. It took many meetings to convince the participants (a) that databases were only one type of informatics resources, and (b) that informatics should focus on a broader range of functions than simply creating, integrating, and providing access to the databases.

3.2 Oracle™ and Google™ Mindsets

Related to equating informatics with databases, there was a strong tendency to think exclusively in terms of structured databases: simple flat Excel files and more complex relational databases. This could be called an 'Oracle™' mindset to distinguish it from the 'Google™' mindset – storage and search based on unstructured repositories and documents. Part of the bias appears to be due to lack of knowledge of today's information systems' ability to index, search, and access large volumes of unstructured data in documents and other sources. Even though researchers used Google™ and similar search engines, few have knowledge of how search engines work. The bias was compounded by an unrealistic equating of the cost of shrink-wrapped databases with the cost of building a structured database for unstructured data.

3.3 Thinking Outside the SiloS

Often, informatics requirements presented by the other core groups involved in the CTS effort appear to be focused on research within their silos, and not across silos. They do not seem to be focused on the truly translational processes of feed-forward and feed-back, but focused on feed-in. While many of the informatics requirements will no doubt facilitate research, it is not clear how they will facilitate translational research. It will perhaps take some more time for the researchers to change their framework through education, consultation, and experience with the proposed CTSII.

3.4 Foundation and Frontier Requirements

A consequence of the issues discussed in the above sections is that most of the requirements presented tend to be what we have called 'foundation' requirements; for example: web-accessible databases and an interactive directory of researchers as a baseline requirement. They are necessary for CTS, and for that matter, any research. They are unlikely, however, to transform the CTS research at the institution and to distinguish one institution from another. That requires the 'frontier' capabilities. These capabilities reflect some of the best practices across different types of organizations and may have to be adapted to CTS. While many of the researchers are aware of these capabilities and have used them, they do not often see their application to CTS. The barriers to the transfer of these best practices are many, one being the inaccurately perceived high cost of performing certain functions. The cost of full-text search is an example of the incorrect perception. The complexity and difficulty of social network analysis is another example of such perception.

It will likely be a while before the researchers will start using CTSII as a tool for CTS research instead of just a service to improve their current research. Until then, the stated requirements are likely to be at the foundation end of the spectrum rather than at the frontier end. Movement of the researchers' thinking along the spectrum will be an important part of the transformation. When, in fact, they move into the frontier, there will likely be a sudden cascade of new and innovative requirements of CTSII.

3.5 Metaphoric Success

The metaphor 'CTSII = Google™ + Facebook™ + Amazon™ + Orbitz™' was very successful in communicating our ideas and to reframe their thinking. Many people, given their age, were not familiar with Facebook™, although some reported their children were using it or a similar system. They were less familiar with Orbitz™ than with Expedia™ and Travelocity™. We chose Orbitz™ as a result of the recent experience of one of the authors with booking a complex international trip. Of course, most researchers were familiar with and had used Google™ and Amazon™.

Almost everybody liked the metaphor once it was explained to them. Many, especially those in informatics related disciplines, intuitively grasped its meaning, and immediately liked it. We still had to explain the application of the metaphor to the design of CTSII, however. The barriers to application of the metaphor were similar to those explained in 3.3 above. The metaphors and the menu have been the anchors of all the recent discussions on CTSII.

3.6 Feed-forward, Feed-backward, and Feed-in

In our eagerness to adhere to the spirit and letter of CTS, we initially proposed feed-forward and feed-back mechanisms as the bases for translation between the four types of research. Feed-in was added in response the expressed need of improving the informatics within each discipline, too; however, the feed-concepts did not appear to resonate with the CTS researchers. As a consequence, we renamed "feed-forward and feed-back" to "two-way

translation of information". We continue to believe, however, that the feed-concepts are central to CTS, and part of the transformation will be to understand and apply these concepts systematically.

In a sense, the CTS paradigm has close parallels to the Continuous Improvement Paradigm that emerged in the context of the Total Quality Management (TQM) movement (Söderholm, 2004). The feed-concepts will likely be as critical to CTS as they have been to Continuous Improvement.

3.7 Technological Bias

Given the focus of most researchers on the technical aspects of informatics – the databases and data warehouses, it took time to incorporate the socio-technical view in the discussions. Clearly, help desk, consultation, education, training, and other user services must be part of the CTSII. The idea of little or no human intervention in this transformation is not realistic.

3.8 Local and Global Transformation

The major impetus for CTS is to transform research. Some of this transformation can be in the CTSII, and some using CTSII. The CTSII transformation can be local or global. Local transformation is one which is innovative within an institution, but not globally. Other institutions may have adopted the innovations earlier. Global transformation is one which is innovative within the institution and without. It is the first of a kind, anywhere.

Both local and global CTSII transformations are necessary for the success of CTS efforts in an institution. Many local transformations may be necessary to bring an institution on par with other institutions; at least a few global transformations will be necessary to provide a competitive advantage to the institution.

4 CONCLUSIONS

An informatics infrastructure for clinical and translational science (CTS) can be complex. We conceptualize the information flows for CTS as a bidirectional flow of feed-forward and feed-back between basic researchers, animal researchers, clinical researchers, and public health researchers, and feed-in among researchers within a discipline. We present a six-dimensional framework as a menu to deconstruct the complexity and help specify the requirements of the clinical and translational science informatics infrastructure (CTSII) for an institution. The framework provides a simple concise way of enumerating all the functions required of a CTSII. Last, we present a case study of our experience in using the framework at our institution. The case study illustrates some of the barriers to the application of the framework and how these barriers can be overcome.

REFERENCES

NIH Office of Communications. NIH News: NIH announces strategy to accelerate medical research progress. September 30, 2003. Available at http://www.nih.gov/news/pr/sep2003/od-30.htm. Accessed May 31, 2007.

NIH Office of Portfolio Analysis and Strategic Initiatives. Re-engineering the clinical research enterprise. April 7, 2006. Available at http://nihroadmap.nih.gov/clinicalresearch/. Accessed May 31, 2007.

NIH Office of Portfolio Analysis and Strategic Initiatives. Re-engineering the clinical research enterprise. March 1, 2007. Available at http://nihroadmap.nih.gov/clinicalresearch/overview-translational.asp Accessed May 31, 2007.

Ramaprasad, A., 1979. The Role of Feedback in Organizational Change: A Review and Redefinition. *Cybernetica*, 22 (2), 105-113.

Ramaprasad, A., 1982. Revolutionary Change and Strategic Management. *Behavioral Science*, 27 (4), 387-392.

Ramaprasad, A., 1983. On the Definition of Feedback, *Behavioral Science*, 28 (1), 4-13.

Ramaprasad, A. and Rai, A., 1996. Envisioning Management of Information. *Omega: The International Journal of Management Science*, 24(2), 179-193.

Söderholm, P., 2004. Continuous Improvements of Complex Technical Systems: a Theoretical Quality Management Framework Supported by Requirements Management and Health Management. *Total Quality Management*, Vol. 15, No. 4, 511–525.

Valenta, A.L., Brooks, I., Laureto, R.A., and Ramaprasad, A., 2007. Informatics to Support Clinical and Translational Science. Journal of Health Information Management (forthcoming).