

TrialX: Using semantic technologies to match patients to relevant clinical trials based on their Personal Health Records

Chintan Patel, Sharib Khan and Karthik Gomadam

Applied Informatics Inc, New York, New York
{chintan, sharib, karthik}@trialx.com

Abstract. Clinical trials form a critical link in the translation of basic biological research into routine clinical practice. However, finding eligible patients for clinical trials is a critical hurdle and a frequent cause for delays in the completion of trials. The lack of comprehensive, efficient, and consumer-friendly online tools has limited consumers ability to pro-actively find the right clinical trials and participate in them. The rising use of Internet as a medium for seeking health information and the advent of online Personal Health Records (PHR) platforms such as Microsoft HealthVault (MHV) and Google Health (GH) have created the opportunity to change the status quo. These platforms are connected to databases of several healthcare organizations and enable patients to import their medical record and utilize it for self-management by using third-party applications built on these platforms. The PHR contains a rich source of information about a patient including their health conditions, medications and laboratory results and opens up the possibility of using this information to provide personalize information recommendations. We have built, TrialX, a consumer-centric tool that matches patients to clinical trials by extracting their PHR information and linking it to the most relevant clinical trials using semantic web technologies. TrialX is currently live and integrated with both MHV and GH and has matched thousands of patients to clinical trials in the last twelve months. We are generalizing the TrialX approach, by creating HealthX, a system to link health information in the PHR to any online health resource.

1 Introduction

Clinical trials are medical research studies performed on animals and humans to evaluate the safety and efficacy of a new drug ¹. They form a critical and essential phase in the lifecycle of drug development and no drugs are approved without completion of successful clinical trials. Every year more than 50,000 clinical trials are conducted in the United States alone. However, finding the right participants (patients or healthy volunteers) has been and continues to be a major bottleneck in the timely completion of these trials. Four out of five

¹ http://en.wikipedia.org/wiki/Clinical_trial. Accessed Oct 03, 2009

trials are delayed and among these, 50% of the trial delays are due to participant recruitment challenges [2, 1]. The situation is particularly troublesome in the domain of oncology, where fewer than 3% of potentially eligible cancer patients enroll in trials [9]. In fact more than 75% of participants are not even aware of them, even though surveys have repeatedly shown that a majority of patients would be open to participating in trials if they had the right information².

From a social and public health perspective, delays in clinical trials have several adverse consequences. One, it entails the loss of human life that could otherwise have benefited from the new intervention (this is specially true with life-altering conditions such as Multiple Myeloma or Cancer, where experimental treatments are the last resort for some patients). The delay in trial completion postpones the entry of a drug in the general market thus postponing its benefits to the large population. Another consequence of the low recruitment rates is that finding patients for trials of uncommon conditions can be challenging and at times impossible, causing the failure of drugs to treat such conditions to make it to the market. The delay has economic consequences too; since a drugs patent period is limited (17 years from the day of filing of the investigational new drug application with the Food and Drug Administration), every days delay in clinical trial results, which on average can take 5-6 years, leads to lost revenues in the range of \$1-\$8 million per day for pharmaceutical companies.³ In total, delays in clinical trials result in over \$10 billion/yr in estimated missed revenues for pharmaceutical companies.

Two fundamental hurdles prevent increased participation of patients in clinical trials. One, the current process of finding eligible participants is largely driven by clinical trial doctors (technically called investigators) or specialized recruitment professionals with limited options for patients to be pro-active in the process of finding and enrolling in trials. Two, the existing participant-oriented tools such as websites that list clinical trials or websites designed to signup patients for clinical trials (such sites are called eRecruitment sites), have several limitations; they lack user-friendliness [8] and many patients dont know enough details of their health conditions (such as what medications they are taking) to find the most relevant trials from the thousands of trials active at any time. Fundamentally new approaches are needed to empower patients with the right tools to help them find the most relevant clinical trials that match their health information without burdening them with complex medical jargon. Two significant trends are converging to create an opportunity for disruption in the current patient recruitment landscape and to make it possible to create the kind of solution described above. One is the rise of Internet as a health information-seeking medium for consumers. The data from the Pew Internet and American Life Project shows that the Internet has become a valued source of health information for a large number of people; these studies estimate that between 75% and 80% of Internet users search for health information online, with a steady

² <http://www.cisr.org/information/facts.asp>

upward trend in this number ³ [7]. On average, 8 million people search for health related information online every day. Websites such as WebMD are visited by millions of health consumers every month. In addition to general health searches, clinical trial information (or information about new treatments) ranks as one of the important health topics that users search for on the Internet. According to the National Library of Medicine, its clinical trials listing website, ClinicalTrials.gov, receives close to 50,000 unique visitors a month and serves more than 40 million page views per month ⁴.

The second converging factor is that the healthcare industry is on the verge of a unprecedented adoption of Healthcare Information Technology (HIT) solutions such as EHRs and PHRs. The recent American Recovery and Reinvestment Act of 2009, has specifically allocated \$34 billion for incentivizing EHR adoption.[10] Riding this thrust in HIT and the consumer health trend, technology corporations including Google and Microsoft are building sophisticated PHR platforms (Google Health ⁵ and Microsoft HealthVault⁶, respectively) to enable consumers to have access to their health information from organizations such as the Cleveland Clinic and be able to use that information for self-management of their health conditions.

The Markle foundations connecting for health collaborative, has defined a PHR as an “electronic application through which individuals can access, manage and share their health information and of others for whom they are authorized, in a private, secure and confidential environment.”[3] The US Health and Human Services and the Office of the National Coordinator for HIT have identified PHRs as a potentially disruptive technology and a national priority [6]. These PHR platforms could be the enablers of a radical paradigm shift in the health information economy; the shift of locus of control of health information to the patient (consumer). A critical differentiating (and potential success) factor of these platforms is their ability to provide open access to software developers to build applications that use information from the PHRs. The central goal of the new PHR platforms, GH and MHV is to let patients use their health information for better self-management by using third-party application services built on these PHRs. Third-party PHR applications can re-use the data within a patients PHR (with due patient consent and with required measures to protect patient privacy) and provide personalized and useful services to the patient. The provision in these PHRs to enable patients to control what they can do with their health information and who they can share the information with, has been described as a tectonic shift in the health information economy by Mandl and Kohane [5]. Such a platform provides an excellent opportunity to use semantic technologies to utilize the information in the PHR to match patients to relevant

³ Pew Internet and American Life Project. Health Information Online. (Available at http://www.pewinternet.org/PPF/r/222/report_display.asp. Last accessed on Nov 2, 2008)

⁴ <http://clinicaltrials.gov/ct2/info/about>

⁵ <http://google.com/health>. Accessed Oct 01, 2009.

⁶ <http://www.healthvault.com>. Accessed Apr 04, 2009

clinical trials. In fact, one of the important PHR use cases described by Mandl and Kohane, the possibility of patients using their PHR information to find clinical trials that match their health profile. If realized, this new PHR model can have a potentially dramatic effect on addressing the current roadblocks that exist in utilizing medical records for research purposes, as patients can now give their consent to third-party application providers such as a clinical trial matching service, as described in this paper. More specifically, we describe details our system TrialX, a consumer-centric application that is built on top of PHRs and enables patients to discover trials that match their health conditions and connects them with the trial investigators.

2 Overview and Information Flow of the TrialX

Figures 1(a) and 1(b) describe the overall information flow within the TrialX system. As Figure 1 illustrates, TrialX is integrated with MHV, GH and Indivo PHRs. These systems in turn are integrated with databases of several healthcare organizations such as hospitals (Cleveland Clinic, New York Presbyterian) and pharmacies (CVS). A patient can import his or her health record from any of the partner organizations of these platforms. After importing, they can choose to allow third-party applications built on these platforms (such as TrialX) to use their record to provide useful services. For example, an application in HealthVault can read the height, weight, age and gender of a patient and calculate their Body Mass Index (BMI) and explain the significance of the same to the patient.

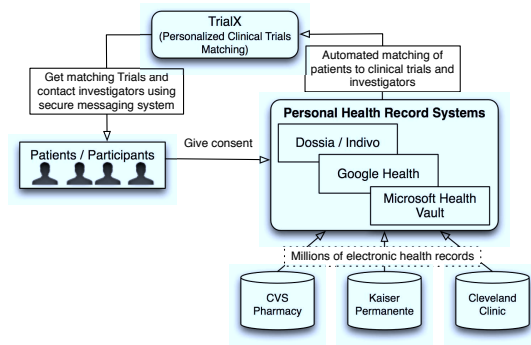
Similarly the patient can add (just as users can add applications on their iPhone or Facebook) the TrialX application to their health record.

After adding the application, the user can generate matching trials by clicking the *show matching trials* link. The TrialX system works in the backend to pull the patients condition, demographics and other information securely and uses this information to generate a list of matching trials on TrialX.com.

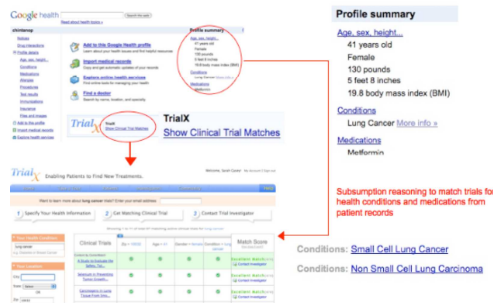
3 Semantic Web in Action: TrialX Under the Hood

Figure 3 illustrates the core components of the TrialX system. The application uses the data from PHRs, along with existing semantic models such as *SNOMEDCT*, *ICD9CM*, and *RxNORM*, to perform three operations:

1. PHR Integration: Integrate the health record information from various PHR providers
2. Conversion of data into Semantic Representation: Convert the data into formal semantic models to enable reasoning and graph based discovery. This task is accomplished by the TripleX component in the system.
3. Matching of Patient Information to Clinical Trial Information: Employ reasoning and graph algorithms for discovering matching clinical trials for a given patient record. The XOperator component is responsible for the semantically expanding the queries, and for data transformation during the



(a)



(b)

Fig. 1. (a) Information Flow between PHRs and TrialX (b) Example: Integration between Google Health and TrialX

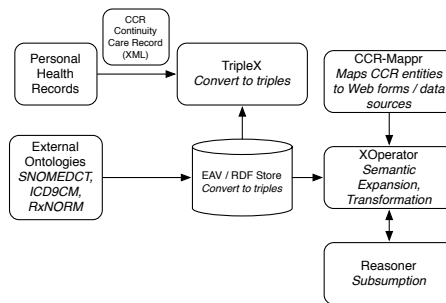


Fig. 2. TrialX: Core components

matching process. We also have employ a reasoner during the matching process.

At the heart of the TrialX platform is our Columbus Matching Technology (CMT). CMT performs clinical record based computational matching of par-

ticipants with clinical trials using semantic and NLP techniques. Using CMT, we analyze key sections of the patient’s health record including demographics, health conditions, laboratory results, medications, and procedures. This information is used in matching the patient with the eligibility criteria information obtained from a trial description. One of the key features of CMT is that the matching is performed at a semantic concept level (biomedical meaning) rather than checking for absence or presence of a criterion at the lexical level. Consider a patient taking the drug Vancomycin. This patient will be matched to a clinical trial that requires patients taking antibiotics. This is because, CMT can infer using semantic conceptual knowledge that Vancomycin is an antibiotic. By using a semantic approach to trial matching, we provide more meaningful results as compared to lexical (term level) search results. The conditions and criteria in the protocols are processed using NLP to identify the most relevant Unified Medical Language System (UMLS) concepts associated with each trial [4]. The UMLS allows one to map different synonyms into a common conceptually and semantically identical concept. For example, using the UMLS one can map terms such as hypertension and high blood pressure to a single concept in the UMLS. When the user enters a query, we dynamically map the user query term into a UMLS concept and retrieve relevant clinical trials indexed with the same concept. This allows consumers to use a wide variety of synonymous terms, which have the same meaning, thereby improving the recall of our query system without compromising accuracy.

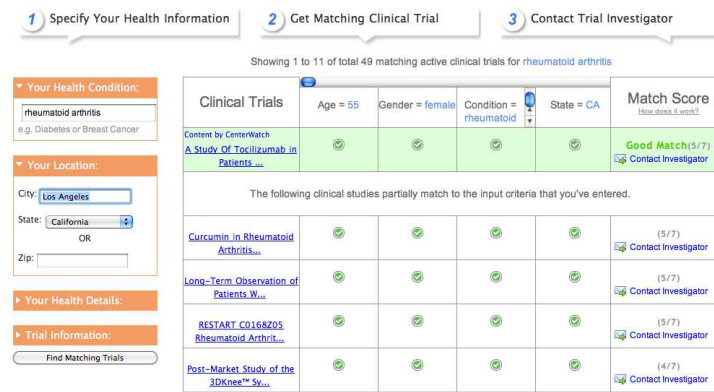


Fig. 3. Search Results Matrix View

Salient to TrialX is the unique matrix view employed in our search results. This view is illustrated in 3. The matrix view, again employs semantic Web principles to identify and highlight the important characteristics that were matched, for each trial that is returned by the search engine. The order of columns in this view is determined by the importance of a particular attribute to the condition for which the user is searching for trials.

In addition to the matching algorithm and search results, TrialX also employs semantic Web techniques to generate form interfaces. When a user (investigator) wants to add a new trial, the system automatically identifies the minimum information required for a match. We use previously mined trial information, along with external ontologies to realize this feature. After identifying the required information, we use the triple store to identify the user interface element (for example: Gender is captured using a radio button, Trial phase is a drop down) and generate the user interface for this trial. This interface generation is also available as a service (<https://trialx.com/widget/>), where investigators can create custom widgets that can be embedded in various Web applications. add a new trial,

4 Discussion and Conclusion

In this paper we have described the need for building online consumer-centric technologies to connect patients to relevant clinical trials. We are leveraging paradigm changing technologies such as PHR in combination with semantic web technologies to match patients based on the information in their record and the information provided for a clinical trial.

The approach utilized to build TrialX provides a mechanism for connecting patients to relevant and personalized information. The system has received enthusiastic response from the users and is actively growing. It provides the framework and the core technologies that can be used to extend the concept a generalized system HealthX, for connecting health consumers to personalized health information resources based on their PHR.

References

1. M. Barrett. erecruiting for clinical trials. 2001.
2. Centerwatch. *An Industry in Evolution*. Centerwatch, 2003.
3. M. Foundation. Connecting americans to their healthcare. *Markle Foundation Connecting for Health*, page 48, 2004.
4. M. A. Lindberg DA, Humphreys BL. he unified medical language system. *T Methods Inf Med.*, 32:281–291, 1993.
5. K. I. Mandl KD. Tectonic shifts in the health information economy. *N Engl J Med.*, 358:1732–1737, 2008.
6. B. D. O. J. S. D. Tang PC, Ash JS. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. *J Am Med Inform Assoc.*, 13:121–126, 2006.
7. P. C. Trusts”. The online health care revolution: How the web helps americans take better care of themselves. 2006.
8. K. S. Valerie M. Online information about cancer clinical trials: Evaluating the web sites of comprehensive cancer centers. *AMIA Annu Symp Proc*, pages 470–474, 2003.
9. R. Winn. Obstacles to the accrual of patients to clinical trials in the community setting. *Semin Oncol*, 21(4):112–117, 1994.